

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

Х А Б А Р Л А Р Ы

ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ
НАУК РЕСПУБЛИКИ КАЗАХСТАН
Қазақстан Республикасының
Ғылым Академиясының
Әл-Фараби атындағы
Қазақ ұлттық университеті

NEWS

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN
al-Farabi Kazakh National University

PHYSICO-MATHEMATICAL SERIES

4 (344)

OCTOBER – DECEMBER 2022

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас директорының м.а. (Алматы, Қазақстан), **Н=5**

РЕДАКЦИЯ АЛҚАСЫ:

КАЛИМОЛДАЕВ Мақсат Нұрәділұлы (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институты бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), **Н=7**

МАМЫРБАЕВ Өркен Жұмажанұлы (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институты директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), **Н=5**

БАЙГУНЧЕКОВ Жұмаділ Жанабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), **Н=3**

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), **Н=23**

СМОЛАРЖ Анджей, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), **Н=17**

ӘМІРҒАЛИЕВ Еділхан Несіпханұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Жасанды интеллект және робототехника зертханасының меңгерушісі (Алматы, Қазақстан), **Н=12**

КИЛАН Әлімхан, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония), ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас ғылыми қызметкері (Алматы, Қазақстан), **Н=6**

ХАЙРОВА Нина, техника ғылымдарының докторы, профессор, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас ғылыми қызметкері (Алматы, Қазақстан), **Н=4**

ОТМАН Мохаммед, PhD, Информатика, коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті (Селангор, Малайзия), **Н=23**

НЫСАНБАЕВА Сауле Еркебұланқызы, техника ғылымдарының докторы, доцент, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының аға ғылыми қызметкері (Алматы, Қазақстан), **Н=3**

БИЯШЕВ Рустам Гакашевич, техника ғылымдарының докторы, профессор, Информатика және басқару мәселелері институты директорының орынбасары, Ақпараттық қауіпсіздік зертханасының меңгерушісі (Қазақстан), **Н=3**

КАПАЛОВА Нұрсұлу Алдажарқызы, техника ғылымдарының кандидаты, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), **Н=3**

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), **Н=5**

МИХАЛЕВИЧ Александр Александрович, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), **Н=2**

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), **Н=42**

«ҚР ҰҒА Хабарлары. Физика-математикалық сериясы».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **№ 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *физика-математикалық сериясы*».

Қазіргі уақытта: «ақпараттық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.

Мерзімділігі: жылына 4 рет.

Тиражы: 300 дана.

Редакцияның мекен-жайы: 050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19

<http://www.physico-mathematical.kz/index.php/en/>

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2022
Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

Главный редактор:

МУТАНОВ Галимкаир Мутанович, доктор технических наук, профессор, академик НАН РК, и.о. генерального директора «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=5**

Редакционная коллегия:

КАЛИМОЛДАЕВ Максат Нурадилович, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), **Н=7**

МАМЫРБАЕВ Оркен Жумажанович, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), **Н=5**

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Саптаева (Алматы, Казахстан), **Н=3**

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), **Н=23**

СМОЛАРЖ Анджей, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), **Н=17**

АМИРГАЛИЕВ Едилхан Несипханович, доктор технических наук, профессор, академик Национальной инженерной академии РК, заведующий лабораторией «Искусственного интеллекта и робототехники» (Алматы, Казахстан), **Н=12**

КЕЙЛАН Алимхан, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=6**

ХАЙРОВА Нина, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=4**

ОТМАН Мохамед, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), **Н=23**

НЫСАНБАЕВА Сауле Еркебулановна, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=3**

БИЯШЕВ Рустам Гакашевич, доктор технических наук, профессор, заместитель директора Института проблем информатики и управления, заведующий лабораторией информационной безопасности (Казахстан), **Н=3**

КАПАЛОВА Нурсулу Алдажаровна, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=3**

КОВАЛЕВ Александр Михайлович, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), **Н=5**

МИХАЛЕВИЧ Александр Александрович, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), **Н=2**

ТИГИНЯНУ Ион Михайлович, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), **Н=42**

«Известия НАН РК. Серия физико-математическая».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **№ 16906-Ж** выданное 14.02.2018 г.

Тематическая направленность: *серия физико-математическая.*

В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*

Периодичность: *4 раз в год.*

Тираж: *300 экземпляров.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

© Национальная академия наук Республики Казахстан, 2022
Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

Chief Editor:

MUTANOV Galimkair Mutanovich, doctor of technical sciences, professor, academician of NAS RK, acting General Director of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H=5**

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), **H = 7**

Mamyrbayev Orken Zhumazhanovich, (Academic Secretary), PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H = 5**

BAIGUNCHEKOV Zhumadil Zhanabaevich, Doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), **H=3**

WOICIK Waldemar, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), **H=23**

SMOLARJ Andrej, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), **H= 17**

AMIRGALIEV Edilkhan Nesipkhanovich, Doctor of Technical Sciences, Professor, Academician of NAS RK, Head of the Laboratory of Artificial Intelligence and Robotics (Almaty, Kazakhstan), **H= 12**

KEILAN Alimkhan, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), **H= 6**

KHAIROVA Nina, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), **H= 4**

OTMAN Mohamed, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), **H= 23**

NYSANBAYEVA Saule Yerkebulanovna, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H= 3**

BIYASHEV Rustam Gakashevich, doctor of technical sciences, professor, Deputy Director of the Institute for Informatics and Management Problems, Head of the Information Security Laboratory (Kazakhstan), **H= 3**

KAPALOVA Nursulu Aldazharovna, Candidate of Technical Sciences, Head of the Laboratory cyber-security, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H=3**

KOVALYOV Alexander Mikhailovich, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), **H=5**

MIKHALEVICH Alexander Alexandrovich, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), **H=2**

TIGHINEANU Ion Mihailovich, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

News of the National Academy of Sciences of the Republic of Kazakhstan.

Physico-matematical series.

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-Ж**, issued 14.02.2018

Thematic scope: *physical-mathematical series.*

Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*

Periodicity: *4 times a year.*

Circulation: *300 copies.*

Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

© National Academy of Sciences of the Republic of Kazakhstan, 2022

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 5-15

<https://doi.org/10.32014/2022.2518-1726.152>

УДК 28.23.29

МРНТИ 28.17.2

А.С. Баймаханова¹, А.Ж. Сейтмуратов^{2*}

¹Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті,
Түркістан, Қазақстан;

²Қорқыт Ата атындағы Қызылорда университеті, Қызылорда, Қазақстан.
E-mail: *angisin_@mail.ru*

DEEP LEARNING АЛГОРИТМІН ҚОЛДАНУ НЕГІЗІНДЕ ЦИФРЛЫҚ ҚҰЖАТТАРДЫ ЖІКТЕУ

Аннотация. Цифрлық құжаттарды сақтауда көптеген ұйымдар үшін құжаттарды жіктеу маңызды процесс болып табылады. Зерттеу барысында Deep Learning алгоритмдерін қолдануды дамыту технологиясы жасалады. Құжаттар алдымен визуалды түрде, онан соң, бірінші кезеңде жіктелмейтін құжаттар табиғи тілде өңдеу (NLP) арқылы жіктеледі. Оның ішінде топтар саны 20 шамасында, ал құжаттар саны 10 000-дай болады. Зерттеулер бір беттік құжаттан бастап жүргізіледі. Бірінші кезеңде құжаттар көрнекі түрде жіктеледі. Екінші кезеңде жіктелмейтін құжаттарды табиғи тілді өңдеу (NLP) арқылы жіктеуге тырысты. Жаттығулар жинағы шамамен 20 классификация арқылы жасалды құжаттар қолмен. Зерттеулер бір бетте жүргізілді құжаттар. Орындаулар жасау арқылы AlexNet көмегімен құжат толық өлшенеді, $\frac{1}{2}$ және $\frac{1}{3}$ бөлігінде тәжірибелер кескін классификациясы үшін, сөздер пакеті және LSTM алгоритмі үшін NLP алынады. Зерттеудің көрсеткіштері сәйкес өлшенді әртүрлі опцияларға жіктеледі. Жетістік деңгейі 75% және 85% болғанда, тиісінше бірінші және екінші кезеңдердің аяқталуы қажет. AlexNet архитектурасы өзгертілгенде жіктеуге келмейтін құжаттардағы кейбір мәтіндерді қолмен енгізуімізге тура келеді. Зерттеу жұмысында алынған нәтижелер, CNN-де құжат тобы белгілі бір ықтималдықпен анықталады. Басқа сөздер менайтқанда, ықтималдықтың белгілі бір мәнін қанағаттандыратын құжаттар тобы қажет. Тақырыптың өзектілігі цифрлық мұрағаттарда көптеген құжаттарды сақтау ұйымдары үшін құжаттарды жіктеу маңызды процесс болып табылады. Жаңартылған AlexNet көмегімен құжат кескінінің

классификациясы AlexNet архитектурасында сегіз зерттеуден қабаттардан тұрады, бес конволюционды және оған үш толық қосылады. Сонымен қатар жұмыста бастапқы AlexNet түпнұсқасының бірінші кіру деңгейін өзгертіп, ең жақсы көрсеткен нәтижелерді алуғадың мүмкін екенін көрсеттік .

Түйін сөздер: Deep Learning, AlexNet, NLP, CNN, цифрлық құжаттар.

А.С. Баймаханова¹, А.Ж. Сейтмуратов^{2*}

¹Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, Туркистан, Казахстан;

²Кызылординский университет имени Коркыт Ата,
Кызылорда, Казахстан.
E-mail: *angisin_@mail.ru*

КЛАССИФИКАЦИЯ ЦИФРОВЫХ ДОКУМЕНТОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ АЛГОРИТМА DEEP LEARNING

Аннотация. Классификация документов является важным процессом для многих организаций при хранении цифровых документов. В ходе исследования создается технология отработки использования алгоритмов глубокого обучения. Документы сначала классифицируются визуально, затем документы, не классифицированные на первом этапе, классифицируются с использованием обработки естественного языка (NLP). Среди них количество групп около 20, а количество документов около 10 000. Исследования ведутся, начиная с одностраничного документа. На первом этапе документы классифицируются визуально, на втором этапе неклассифицируемые документы пытались классифицировать с помощью обработки естественного языка (NLP). Комплект упражнений был создан с использованием около 20 классификационных документов вручную. Исследования проводились на одностраничных документах. Выполняя исполнения, документ полностью измеряется с помощью AlexNet, в 1/2 и 1/3 экспериментов получается NLP для классификации изображений, пакета слов и алгоритма LSTM. Показатели исследования классифицируются по разным вариантам, которые соответственно измеряются. При успешности 75% и 85% должны быть выполнены первый и второй этапы соответственно. При изменении архитектуры AlexNet нам приходится вручную вводить некоторый текст в документы, которые невозможно классифицировать. По результатам, полученным в исследовательской работе, в CNN группа документов определяется с определенной вероятностью. Другими словами, нам нужна группа документов, удовлетворяющих определенному значению вероятности. Актуальность темы в цифровых архивах классификация документов является важным процессом для многих организаций по хранению документов. Классификация

изображений документов с обновленной AlexNet состоит из восьми исследовательских слоев в архитектуре AlexNet, пяти сверточных и трех полностью аддитивных. Кроме того, в работе мы показали, что можно изменить первый уровень доступа оригинального AlexNet и получить наилучшие результаты.

Ключевые слова: глубокое обучение, AlexNet, NLP, CNN, цифровые документы.

A. Baimakhanova¹, A. Seitmuratov^{2*}

¹Khoja Akhmet Yassawi International Kazakh-Turkish University,
Turkistan, Kazakhstan;

²Korkyt Ata Kyzylorda State University, Kyzylorda, Kazakhstan.
E-mail: *angisin_@mail.ru*

CLASSIFICATION OF DIGITAL DOCUMENTS USING DEEP LEARNING ALGORITHM

Abstract. Document classification is an important process for many organizations in digital document storage. In the course of the research, the technology of developing the use of Deep Learning algorithms is created. The documents are first classified visually, then the documents that are not classified in the first stage are classified using natural language processing (NLP). Among them, the number of groups is about 20, and the number of documents is about 10,000. Research is conducted starting from a one-page document. In the first stage, documents are visually classified. In the second stage, unclassifiable documents were tried to be classified using natural language processing (NLP). A set of exercises was created using about 20 classification documents manually. Studies were conducted on one page documents. By making the executions, the document is fully measured using AlexNet, in $\frac{1}{2}$ and $\frac{1}{3}$ of the experiments, NLP is obtained for image classification, word pack and LSTM algorithm. The indicators of the study are classified into different options, which are measured accordingly. With a success rate of 75% and 85%, the first and second stages should be completed respectively. When changing the architecture of AlexNet, we have to manually enter some text into documents that cannot be classified. According to the results obtained in the research work, in CNN a group of documents is determined with a certain probability. In other words, we need a group of documents that satisfy a certain probability value. Relevance of the topic In digital archives, document classification is an important process for many document preservation organizations. The updated AlexNet document image classification consists of eight exploratory layers in the AlexNet architecture, five convolutional layers, and three fully additive layers. In addition, in the work we have shown that it is possible to change the first access level of the original AlexNet and get the best results.

Key words: Deep Learning, AlexNet, NLP, CNN, digital documents.

Кіріспе. Қазақстан Республикасы Үкіметінің 2017 жылғы 12 желтоқсанда №827 Қаулысымен «Цифрлық Қазақстан» мемлекеттік бағдарламасы бекітілді. Жолдауа еліміздің ұзақ мерзімді мақсаттарды және аз уақыт ішіндешілуі қажет міндеттер нақты анықталған. Сол міндеттердің ішіндегі елімізде IT-саласындамыту мәселесі туралы: «Біз цифрлық технологияларға негізделген болашағызор бағыттарға айрықша мән беретін боламыз» - деп, IT-саласын дамытуды ерекше бақылауда ұстауымыз қажеттілігін атап айтты. («Цифрлық Қазақстан»-2019). Атап айтқанда, медицина қаржы, білім және т.б салаларда кеңінен қолданыс тауып, халықтың әлеуметтік жағдайын жақсартуға септігін тигізуде. Цифрландыру процесін жүзеге асыру еліміздің бәсекеге қабілетті елдер қатарына енуіне мүмкіндік береді. Бұл процесті жүзеге асыру мақсатында бес бағыт бойынша жұмыс жасау қабылданды: экономикалық салаларды сандық түрге келтіру, сандық мемлекетке көшу, сандық жібек жолын іске асыру, адами капиталды дамыту, инновациялық экосистемелерді құру (Гурянова, 2017).

Бағыттардағы цифрлық деректердің үлкен көлемі және олардың құрылымданбауы ақпараттық кедергілер мен мәселелер тудырады. Сондықтан да осы мәселелерді шешуде жаңа шешімдерді, жаңа тәсілдерді, сонымен қатар ақпараттарды өңдеуші жаңа алгоритмдер қалыптастыру қажет.

Құжаттарды жіктеу кеңседе маңызды рөл атқаратын жұмыс, олар алдынала анықталған құжаттарды жіктеуге арналған және цифрлық мұрағатта сақтайды.

Кескіндерді жіктеу Deep Learning алгоритмдері арқылы жасалады және оның тиімділігі AlexNet-пен салыстырылады. Нәтижесінде NLP жіктелуіндегі сөз қоры мен LSTM алгоритмі әзірленеді.

Тақырыптың өзектілігі цифрлық мұрағаттарда көптеген құжаттарды сақтау ұйымдары үшін құжаттарды жіктеу маңызды процесс болып табылады. Зерттеу барысында Deep Learning алгоритмдерін қолдануды дамыту технологиясы жасалады. Құжаттар алдымен визуалды түрде, онан соң, бірінші кезеңде жіктелмейтін құжаттар табиғи тілде өңделінеді (NLP) арқылы жіктеледі. Оның ішінде топтар саны 20 шамасында, ал құжаттар саны 10 000-дай болады. Зерттеулер бір беттік құжаттан бастап жүргізіледі. Кескіндерді жіктеу Deep Learning алгоритмдері арқылы жасалады және оның тиімділігі AlexNet-пен салыстырылады. Нәтижесінде NLP жіктелуіндегі сөз қоры мен LSTM алгоритмі әзірленеді.

Классификация құжаттардың қазіргі таңда маңызды жұмыс атқарады. Құжаттарды алдын ала анықталғандарға жіктеуге бағытталған топтарды және оларды сәйкесінше цифрлық мұрағатта сақтау қажет. Жіктеуге бағытталған құжаттарды алдын-ала жіктелген топтарда сақталады. Бұл жұмыста енгізілген зерттеу 20 топқа бөлініп, соның ішінде 10 000-ға жуық құжат құрастырылады. Берілген құжаттар цифрлық мұрағатта сақталады және қажет болған жағдайда мұрағаттан алынып тасталады. Біздің жұмысымыз үшін әзірленген әдістер мен алгоритмдер құжаттарымызды кескін бойынша

жіктеленіледі. Терең оқыту, яғни терең құрылымдық оқыту – бұл жасанды нейрондық желілерге негізделген машиналық оқыту әдістерінің кең тобының бөлігі болып табылады. Оқытуды бақылауға, ішінара бақылауға немесе бақылауға болмайды (Bengio т.б., 2013).

Дереккөздерне сүйенсек, Фрэнк Розенблатт заманауи терең оқыту жүйелерінің барлық негізгі компоненттерін жасап, зерттегенін көрсетеді (Тапперт т.б., 2019). Интеллектуалды ақпараттық жүйелер тұрғысынан қарағанда бейне – бұл деректерді талдау жасаушы деректер жиынтығынан ерекшеліп алуға және қарастырылып отырған есептің шарттарымен сәйкес басқа объектілер мен топтауға мүмкіндік беретін нақты немесе абстрактілі объектілердегі (процестердегі, құбылыстардағы) деректер жиынтығы. Бастапқыда алгоритм мен әдіс таңдалынады аймағымызға сәйкес қолдану барысында оң нәтижеге қол жеткізіледі. Алдымен олар құжаттардың ерекшеліктерін анықтауға тырысты, содан кейін құжат топтарда зерттелді. Екіншіден, қолдануға болатын алгоритмдер мен әдістер құжатты сурет ретінде жіктеу зерттелді. Бұл зерттеудегі құжаттар, әдетте, мәтіннен жәнестелермен терілген бір беттен тұрады. Құжаттарымызды анықтау, талдау жасау, құжаттарды жіктеу үшін әртүрлі алгоритмдер мен әдістер қолданылды. Нәтижелеріміз AlexNet алгоритмі арқылы алынған.

Қол жеткізілген жетістік AlexNet алгоритмінің көмегімен шамамен 75% құрайды. Табиғи тілөндеу әдістері кейінірек топқа кірмеген адамдар үшін сыналды, AlexNet алгоритмін қолдана отырып анықтай алдық. Бұл тұрғыда алгоритм «Сөздік қорымыз» өте сәтті болды. Осы екі деңгейлі зерттеудің соңында жалпы өнімділік шамамен 95% көрсеткішке жеткізу болды. Басқалары жіктеуге қарапайым, бір беттік мәтіндер мәтіндерді жіктеуге негізделген. Терең конвульсиялық нейрондық желілер CNN кескінді классификациялауда кеңінен қолданылуда. AlexNet моделі кескіндерді классификациялауда терең оқытуда қолданыла бастады. Конвульсиялық нейрондық желілер CNN параметрлердің санын анықтайтын және конверсия мәселелерін тиімді шешеді. Терең оқытуға негізделіп отырған модельдер мәтіндерімізді зерттеу барысында көптеген машиналық оқыту әдістерден басым болуда.

Мәтіндерімізді, бейнемізді жіктеуге байланысты негізгі мәселе-оқытудың ең қолайлы түрін анықтау өте маңызды. Осы оқу бастамаларының көпшілігі өнімділікті арттыруға, қателерді азайтуға және қателіктердің алдын алуға көмектеседі. Алайда жұмысты жіктеудің негізгі түрлері мен әдістерін бөліп көрсету арқылы тану жұмысымыз жүргізеді (Mohammed т.б., 2021). Қазіргі таңда қауіпті шабуылдардың басым бөлігі терең білім негізінде яғни, кескін классификаторларын алдауға бағытталуда. Суреттерді классификациялау барысында шабуылдарды қауіпсіздікке сезімтал ақпараттық жүйеге қарсы қандай шабуылдар жасалуы мүмкін деген болжамдарды анықтайтын әртүрлі қауіп-қатер модельдеріне жатқызуға болады (Machado т.б.-2021:38).

Классификациялаудың тағы бір кең таралған түрі ол конвульсиялық нейрондық желі моделі болып табылады (Minaee т.б.2019). Соңғы жылдары

CNN моделі компьютерлік көру және табиғи тілді өңдеу міндеттерінің кең ауқымды тиімдісі болды. Ұсынылған құрылымды бағалау барысында негізгі модельдерді оқыту үшін қолданылатын сәулет түрі LSTM және CNN архитектурасының әлсіз байланысы LSTM-CNN деп те аталады, ол LSTM және CNN екеуін де біріктіреді (Girshick т.б., 2014).

Материалдар мен әдістер. Технологияның үздіксіз даму барысында терең оқыту әдісі медициналық кескіндерді жіктеуде дәстүрлі машиналық оқыту мәселелерін шешудің жаңа түрін ұсынатын медициналық кескіндерді талдаудың тандаулы әдісіне айналған. Терең оқыту әдісі компьютерлік көру, содан кейін медициналық кескіндерді көруде кеңінен қолданылады. Терең оқытуда және оның құрылымында—оқытушысыз оқыту, бақыланатын оқыту және т.б. Жіктеудің ешқандай нұсқауынсыз CNN-дің дәстүрлі моделін үйрету қиын. Сондықтан, тақырып саласы білімге және тиісті сарапшылардың тәжірибесіне сүйенеді, CNN-ге негізделген қолданыстағы әдістердің көпшілігі MRI диагностикалық моделін құру үшін ақпараттық аймақты (мысалы, гиппокамп немесе учаске) эмпирикалық түрде анықтайды, бұл ми ауруларында диагностикалауда кезінде терең нейрондық желінің тиімділігін төмендетуі мүмкін. Дәстүрлі әдістермен салыстырғанда, терең оқыту модельдері жіктеу үшін суреттердің ерекшеліктерін автоматты түрде зерттеуге болады, диагностикалық өнімділік артады. Терең оқыту технологиясы жарнаманы дәл анықтау үшін қосымшаларды қолдануды қажет етеді.

Харли және т.б. (Harley т.б., 2015) берілген құжаттардың кескіні үшін терең оқыту мүмкіндіктерді қолдана отырып, құжаттардың кескіндерін жіктеу және шығару үшін CNN қолданылады. CNN функциялары сығылуға төзімді әрі құжаттардағы суреттерді оқиды. CNN жақсы тасымалданады әрі құжаттарды талдау тапсырмаларын орындау барысында орындауда белгілі бір мүмкіндікті оқытуға жеткілікті. CNN көп суретті өндеудің басқа әдістеріне қарағанда сәтті көрсеткіштерге ие, CNN байланысты емес суреттері қолданылған құжаттарға қатысты тапсырмаларды өте жақсы орындайды. Афзал және т.б. (Afzal т.б., 2015) CNN-ге негізделген құжаттарды жіктеуді зерттеуде дайындалған мәліметтердің өте көп мөлшері қажет деді.

Зерттеулерде тек жіктеу дәлдігіне қол жеткізу үшін бір сыныпқа 20 топтық үлгісі көрсетіледі. Олар үйренген AlexNet Imagenet суреттері олардың жұмысында қолданады. ImageNet-бұл мәліметтер жиынтығы 500-ден астам суреттер қамтылады, дайындалған әр түйінде және жалпы суреттер қарастырылады. AlexNet-бұл құрамында суреттер 1000 санат. ImageNet құрамында әр түрлі бұл зерттеу құжаттардың 10 түрін зерттеді. Құжаттардың бұл түрлері: дипломдық жұмыстар, бұйрықтар, транзакциялар, шығыс, кіріс журналдары, хаттамалар, ведомства, аттестаттар т.б. көрсетілген. CNN-нің салалардағы зерттеулер басқа зерттеулеріден өзгеше болып табылады. Deep Learning негізгі цифрлық құралдардың көмегімен құжаттарды ретпен басқаруды қамтамасыз етеді. CNN әдісі бейнені тану барысында құжат кескініндегі ең сәтті процесті жіктеу әдісі болып табылады.

Кадам бойынша орындалатын кезеңдер:

-Құжаттарды классификациялау жоспарланған, алдын-ала құрастырылған топтарды қарастырады. Жинақталған бейнелерді жіктейміз. Олар латын немесе кириллица тілінде болуы мүмкін.

- Терең оқыту әдістерін қолданатын қадам үшін AlexNet-ті таңдаймыз, себебі ол суретті тану немесе жіктеу үшін өте тиімді. Егер, біз мәтінді жіктей алмасақ немесе терең оқыту алгоритмдерін қолдана алмасақ, біз NLP-ді қолдана аламыз.

-10 000 құжатты қолмен жіктей отырып жинау қажет, бірақ алдымен топ сөзін таңдаудан бұрын топтың мәтіндік құжаттарын шешеміз. AlexNet RGB кескіндерінде оқытылады және шығарылады.

Үлкен және терең CNN классификациялауда Imagenet-те 10000-ға жуық жоғары ажыратылған суреттерді 20 түрлі топтарға бөлу үшін жаңа әдіс қолданылады. Жоғары оңтайландырылған графикалық суреттер іске асыруды жазуда іске асырылады (Krizhevsky т.б., 2017).

Терең оқыту CNN әдісін процестердің ең сәтті әдісі құжаттардың кескіндерін жіктеу барысында орындалады. Олар бірнеше факторларды қарастырды. Құжаттарды өңдеу кезінде CNN өнімділігі аса зор әсер етеді. Олароқу және оқу кезінде ауысымдық түрлендірулерді қолдануда үлкен кіріс кескіндерін пайдалануда үлкен өсуге әкелетін өнімділік болып табылған (Tensmeyer т.б., 2017). Классификация тапсырмасының қолданыстағы алгоритмдері стратегиялары құжаттар кескіні қатені екі есе азайтуға мүмкіндік береді (Afzal т.б., 2017). Санның әсерін талдауоқу құжатының суреттері және жіктеу мүмкіндіктеріне арналған басқа параметрлер: дипломдық жұмыстар, бұйрықтар, транзакциялар, шығыс, кіріс журналдары, хаттамалар, ведомства, аттестаттар т.б. бойынша жұмыс істеді. AlexNet-те жұмыс істеу барысында көптеген мәселелер туындауы мүмкін.

Ал, жаттығу деректерін Alex Net-ке Python бағдарламасында қолдануға болады. AlexNet-ті оқытқаннан кейін оны ұзақ жіктеу құжаттарын қолдану үшін пайдалануға болады. Alex Net қолданбасын қолдану үшін бір ғана үлгі алынады және бұл құжаттың тобы–міндетті түрде белгілі болады. Егер, AlexNet құжаттар тобын белгілесек болғаны, жіктей алмаған жағдай қарастырылады. Біз OCR Optical Character Recognition арқылы мәтінге түрлендіруге кірістік. Сөз таптарын тауып қолмен енгізуге болады. Талдау жүргізу барысында суреттер де, мәтін де құжаттар және нәтиже алуда екі талдаудың нәтижелерін біріктіре отырып құжаттар кескіндерінің классификациялау, бейнелеу әдісімен мәтінді тану нәтижелерін талдау барысында тиімдірек болады (Jain т.б., 2019)

Құжаттарды жіктеу мәселелеріне өндіріс ортасы мен құжаттарды жіктеудің қолданыстағы әдістерінде назар аудару қажет. Екі кезең ұсынылады: бірінші кезеңде, тереңдетілген желілік оқыту. Бұл мүмкіндіктерді алу тәсілі ретінде жұмыс істейді. Екінші кезеңде, тәуелсіз білім беру жүйесі, яғни CNN және LM нәтижелерін салыстыра отырып, қорытынды шығарамыз. Әлемге негізделген

құжаттарды құру кезінде оқудың жаңа перспективалары нақты уақыттағы ауқымды қосымшаларға сәйкес жіктелуде (Kolsch т.б., 2017). Жаңа жіктеу әдісін күрделі геометриялық құжаттарға ұсынамыз. Олардың көзқарасы негізделген нысандарын жылдам және сенімді анықтау алгоритмдеріне тығыз байланысты.

Ұсынылған әдістің артықшылықтары:

- жоғары жылдамдық;
- автоматты түрде құру мүмкіндігі;
- өңделмеген бейнелерді алдын ала өңдеудің кез-келген кезеңдерінде өңдеу.

Жұмысымызда ұсынылған алгоритмдер құжаттарды классификациялауға ғана емес, сонымен бірге құжаттарды анықтауға да мүмкіндік тудырады және суреттегі құжаттардың орналасуын және бағытын көрсетеді (Usilin т.б., 2010).

CNN мәселелері көптеген шығарылымдарға және ең тиімді параметрлерге ие болады. Олардың ұсынысы бойынша жақсартылған AlexNet негізінде суреттерді сұрыптау әдісі жанартылады. Бұл әдіс AlexNet қабаттарын толықтырады, толық қабатқа сілтемелер бойынша сұрыпталады.

Нәтижелер. Нәтижесінде ұсынылған әдісте параметрлер саны және арақатынас параметрлер пропорциясына толық байланыс деңгейі де жоғарылады сонымен қатар, AlexNet-пен салыстырғанда жіктеу дәлдігін арттырады (Li т.б., 2020).

Жалпы жұмыстарды бағаланған кезде, CNN құжатты терең жинақтауда, кескіндеуде, өңдеудегі ең табысты әдіс болып табылады. CNN-дегі өзгерістер және осы топтамадағы AlexNet-ті қарастырсақ.

Зерттеуіміздің бірінші кезеңінде, Alexnet көмегімен құжаттарды жіктеуге тырыстық және өнімділігін өлшедік. Өзгертулер енгізу арқылы табысты нәтижелерге қол жеткізу үшін AlexNet-ті толық мүмкіндігін қолдандық. Келесі қадамымыз құжаттарды жіктеуге тырыстық, мәтіндік классификатор арқылы суреттерді өңдеу арқылы жіктеуге болатын және жіктеуге болмайтын әдістерді қарастырдық. Құжаттардың суреттерін AlexNet көмегімен жіктеу, суреттерді жіктеу әдісімен құжаттарды жіктеу үшін 10000 құжаттан тұратын оқыту мәліметтер жиынтығы дайындалды (Сурет 1). Құжаттардың 20-ға жуық топтары ажыратылады.



Сурет 1 – Alexnet көмегімен құжаттарды жіктеу

Университеттің архив бөліміндегі құжаттар келесі түрлерде жіктеледі:

- 1) оқудан шығарылған немесе бітірушілердің жеке құжаттары;
- 2) университет қызметкерлерінің жұмыстан босатылғаннан кейін жеке құжаттары;
- 3) дипломдық жұмыстар (жобалар);
- 4) студенттер құрамына қабылдау бұйрықтары;
- 5) жеке құжаттарды қабылдау туралы актілер;
- 6) есепке алу құжаттары;
- 7) жалпы іс жүргізу: кіріс және шығыс құжаттары;
- 8) СМК жазбалары;
- 9) кадрлық есепке алуды жүргізу;
- 10) мемлекеттік органдардан студенттің университетте оқу фактісін растау туралы сұраныстар;
- 11) білім алушылар ауысқан жағдайда басқа білім беру ұйымдарының сұраныстары.

12) іс жүргізуде толтырылған басқа да құжаттар т.б. бойынша жұмыс істелінді. AlexNet-тің сәттілігін бағалау үшін 10000 құжат тексерілді

Сканерлеу өнімділігін өлшеу үшін әртүрлі өлшемдер сыналды. Мысалы, 75x75 пиксель және 200x200 пиксель, ақ-қара және 256 сұр деңгейдегі. Сканерлеу кезінде 75x75 пиксель ажыратымында кара-ақ форматта сканерлеу құжаттардың жеткілікті оқылмайтындығы анықталды.

Құжаттың класын анықтау үшін беттің жоғарғы жағын зерттеу жеткілікті болады. Осы себепті толық бет, жоғарғы жартысыбет және жоғарғы 1/3 беттер сканерленді және өнімділіктің әсерін өлшеуге әрекет жасалды.

Талқылау. Құжаттарды жіктеу оның кескіндерін жіктеу әдісіне негізделген әдіс осы зерттеу аясында жасалады. Құжаттардың суреттерін жіктеу AlexNet негіздері ретінде пайдаланылды, бірақ өзгертілген. Сондай-ақ, сканерлеу жеткілікті екендігі анықталған. Толық беттің орнына жоғарғы 1/3 бет өлшемі тура келеді. Бұл шешімсканерлеу процесін жеделдету тұрғысынан маңызды. LSTM әдістерін сурет бойынша жіктелуі мүмкін емес құжаттароларды өңдеу әдісін қолмен енгізу әдістермен жіктеуде қолдануға болады. Осы кезеңдердің соңында көптеген топтарға жіктелегенқұжаттардың 99% - ы сақталады, ал қалған 1%-ы қолмен енгізілуі мүмкін. Барлық сканерленген құжаттар қара және ақ түсті болады.

Алайда, мұның бәрін жіктеуге болады, ал сканерлейтін құжаттарда 256 сұр деңгейі бар. AlexNet архитектурасы өзгертілгенде жіктеуге келмейтін құжаттардағы кейбір мәтіндерді қолмен енгізуімізге тура келеді.Зерттеу жұмысындаалынғаннәтижелер, CNN-деқұжаттобы белгілі бірықтималдықпен анықталады. Басқа сөздермен айтқанда, ықтималдықтың белгілі бір мәнін қанағаттандыратын құжаттар тобы қажет. Жаңартылған AlexNet көмегімен құжат кескінінің классификациясы AlexNet архитектурасында сегіз зерттеуден қабаттардан тұрады, бес конволюционды және оған үш толық қосылады. Біз тек бастапқы AlexNet түпнұсқасының бірінші кіру деңгейін өзгертіп, ең жақсы көрсеткен нәтижелерді алуға мүмкіндік береді.

Information about authors:

Baimakhanova Aigerim – Doctoral student, Khoja Akhmet Yassawi International Kazakh-Turkish University *aigerimka01@mail.ru*. <https://orcid.org/0000-0002-5364-0146>;

Seitmuratov Angisin – Doktor of Physical and Matematical Sciences, Professor, The Korkyt Ata Kyzylorda State University. Kyzylorda. Kazakhstan. *angisin_@mail.ru*. <https://orcid.org/0000-0002-9622-9584>.

ӘДЕБИЕТТЕР:

A.W. Harley, A. Ufkes, and K.G. Derpanis “Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval,” 13th IAPR International Conference on Document Analysis and Recognition, 2015.

A. Krizhevsky, I. Sutskever, G.E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Communication of the ACM, Volume 60, Issue 6, June 2017 pp 84-90.

A. Kolsch, M.Z. Afzal, M. Ebbecke, M. Liwicki, “Real-Time Document Image Classification using Deep CNN and Extreme Learning Machines”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

Bengio Y.; Courville A.; Винсент П. (2013). «Репрезентативное обучение: обзор и новые перспективы». IEEE Transactions по анализу шаблонов и машинному анализу. 35(8): 1798–1828. *arXiv: 1206.5538* . DOI:10.1109 / *tpami*.2013.50. PMID 23787338. S2CID 393948.

Гурянова (Кибитова) В.Н. Ансамбль алгоритмов для определения ишемической болезни сердца //Сборник тезисовXXVI международный научной конференции Ломоносов.– М:Издательский отдел факультета вычислительной математики и кибернетики МГУ имени М.В. Ломоносова,2017.–С.15-17.

Girshick R., Donahue J., Darrell T. & Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

Mohammed A. & Kora R. (2021). An effective ensemble deep learning framework for text classification. Journal of King Saud University-Computer and Information Sciences.

Machado G.R., Silva E. & Goldschmidt R.R. (2021). Adversarial Machine Learning in Image Classification: A Survey Toward the Defender’s Perspective. ACM Computing Surveys (CSUR), 55(1), 1-38.

Minaee S., Azimi E. & Abdolrashidi A. (2019). Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. arXiv preprint arXiv:1904.04206.

M.Z. Afzal, S. Capobianco, M.I. Malik, S. Marinai, T.M. Breuel, A. Dengel, and M. Liwicki, “DeepDocClassifier: Document Classification with Deep Convolutional Neural Network”, 13th International Conference on Document Analysis and Recognition (ICDAR), 2015.

M.Z. Afzal, A. Kölsch, S. Ahmed, M. Liwicki, “Cutting the Error by Half: Investigation of Very Deep CNN and Advanced Training Strategies for Document Image Classification”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

R. Jain and C. Wigington, “Multimodal Document Image Classification”, 15th IAPR International Conference on Document Analysis and Recognition, 2019.

C. Tensmeyer and T. Martinez, “Analysis of Convolutional Neural Networks for Document Image Classification”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

S. Usilin, D. Nikolaev, V. Postnikov and G.Schaefer, “Visual Appearance Based Document Image Classification”, Proceedings of IEEE 17th International Conference on Image Processing. 2010.

S. Li, L. Wang, J. Li and Y. Yao, “Image Classification Algorithm Based on Improved AlexNet”, MBIDAS 2020 Journal of Physics: Conference Series, 2020.

Тапперт, Чарльз С. (2019). «Кто отец глубокого обучения?» . 2019 Международная кон-

ференция по вычислительным наукам и вычислительного интеллекта (CSCI). IEEE. С. 343–348. DOI: 10.1109/CSCI49370.2019.00067. ISBN978-1-7281-5584-5. S2CID 216043128. Проверено 31 мая 2021 года.

«Цифрлық Қазақстан» мемлекеттік бағдарламасы // (алынған күні 13.05.2019). https://prime-minister.kz/kz/page/view/tsifrluk_kazakstan_memlekettik_bagdarlama

REFERENCES:

A.W. Harley, A. Ufkes, and K.G. Derpanis “Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval,” 13th IAPR International Conference on Document Analysis and Recognition, 2015.

A. Krizhevsky, I. Sutskever, G.E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Communication of the ACM, Volume 60, Issue 6, June 2017 pp 84-90.

A. Kolsch, M.Z. Afzal, M. Ebbecke, M. Liwicki, “Real-Time Document Image Classification using Deep CNN and Extreme Learning Machines”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

Bengio Y.; Courville A.; Винсент П. (2013). «Репрезентативное обучение: обзор и новые перспективы». IEEE Transactions по анализу шаблонов и машинному.

Guryanova (Kibitova) V.N. An ensemble of algorithms for the definition of ischemic heart disease//Sbornik tezisov XXVI mezhdunarodnyy nauchnoy konferencii Lomonosov.–M: Izdatelskiy otdel fakulteta vychislitelnoy matematiki i cybernetiki MGU imeni M.V. Lomonosova, 2017.–P. 15-17.

Girshick R., Donahue J., Darrell T. & Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

Mohammed A. & Kora R. (2021). An effective ensemble deep learning framework for text classification. Journal of King Saud University-Computer and Information Sciences.

Machado G.R., Silva E. & Goldschmidt R.R. (2021). Adversarial Machine Learning in Image Classification: A Survey Toward the Defender’s Perspective. ACM Computing Surveys (CSUR), 55(1), 1-38.

Minaee S., Azimi E. & Abdolrashidi A. (2019). Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. arXiv preprint arXiv:1904.04206.

M.Z. Afzal, S. Capobianco, M.I. Malik, S. Marinai, T.M. Breuel, A. Dengel, and M. Liwicki, “DeepDocClassifier: Document Classification with Deep Convolutional Neural Network”, 13th International Conference on Document Analysis and Recognition (ICDAR), 2015.

M.Z. Afzal, A. Kölsch, S. Ahmed, M. Liwicki, “Cutting the Error by Half: Investigation of Very Deep CNN and Advanced Training Strategies for Document Image Classification”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

R. Jain and C. Wigington, “Multimodal Document Image Classification”, 15th IAPR International Conference on Document Analysis and Recognition, 2019.

C. Tensmeyer and T. Martinez, “Analysis of Convolutional Neural Networks for Document Image Classification”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

S. Usilin, D. Nikolaev, V. Postnikov and G. Schaefer, “Visual Appearance Based Document Image Classification”, Proceedings of IEEE 17th International Conference on Image Processing. 2010.

S. Li, L. Wang, J. Li and Y. Yao, “Image Classification Algorithm Based on Improved AlexNet”, MBDA 2020 Journal of Physics: Conference Series, 2020.

Tappert, Charles S. (2019). “Who is the father of deep learning?”. 2019 International Conference on Computational Sciences and Computational Intelligence (CSCI). IEEE. pp. 343-348. DOI : 10.1109 / CSCI49370.2019.00067 . ISBN 978-1-7281-5584-5. S2CID 216043128 . Verified on May 31 , 2021 .

“Digital Kazakhstan” state program / https://primeminister.kz/kz/page/view/tsifrluk_kazakstan_memlekettik_bagdarlamasi (13.05.2019).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 16-29
<https://doi.org/10.32014/2022.2518-1726.153>

УДК 372.851.02., 372.800.4.02

**М.А. Болатбек^{1*}, Ш.Ж. Мусиралиева², К. Багитова^{2,3}, А.Т. Нюсупов²,
Е. Абайұлы²**

¹Пассау университеті, Пасса, Германия;

²Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;

³Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан.

E-mail: bolatbek.milana@gmail.com

ВЕБ-РЕСУРСТАРДАҒЫ ФИШИНГТІК ХАБАРЛАМАЛАР ЖӘНЕ ОЛАРДЫ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ АРҚЫЛЫ АНЫҚТАУ

Аннотация. Қазіргі таңда веб-ресурстардың қолданысының артуына байланысты киберқылмыс түрлері де артып келеді. Солардың ішінде жиі кездесетін қылмыс түрлерінің бірі фишинг болып табылады, фишинг – пайдаланушыларды алдау және оның құпиясөзін, несие картасының нөмірімен қатар басқа да құпия түрдегі ақпараттарды алу әдістерінің жиынтығы. Көбінесе шабуылдаушылар өздерін электрондық поштадағы қызметкер немесе телефон қоңырауларының танымал ұйымдардың өкілі ретінде көрсетеді. Фишинг – кибершабуылдардың ең қарапайым тәсілі, бірақ ең қауіпті және тиімді әдістердің бірі. Фишингтік хабарламалар қарапайым заңды хабарламаларға ұқсас болуы мүмкін, шабуылдаушы Интернет қолданушысына бұрмаланған веб-сайттарды, сілтемелерді ұсыну арқылы өз мақсатына жетуі мүмкін. Қазіргі уақытта фишингке қарсы тәсілдер сарапшылардан фишингтік сайттардың белгілерін анықтауды және фишингтік веб-сайттарды анықтау үшін үшінші тарап қызметтерін пайдалануды талап етеді. Берілген жұмыста фишинг түсінігіне, оның түрлеріне анықтама беріледі. Берілген бағыттағы соңғы ғылыми мақалаларға шолу жасалады. Сонымен қатар авторлар фишингтік мәліметтерді анықтау үшін машиналық оқыту әдістерін оқытуға және тестілеуге арналған фишингтік мәтіндер жинағын құрастырып, аталған жинақ бойынша әр түрлі машиналық оқыту әдістерін қолданған.

Түйін сөздер: веб-ресурс, киберқауіпсіздік, фишинг, мәліметтер жинағы, мәтінді жіктеу.

Жұмыс AP15473408 «Әлеуметтік желілердегі экстремистік мазмұнды анықтау модельдерін және әдістерін құрастыру» жобасының аясында орындалды.

М.А. Болатбек^{1*}, Ш.Ж. Мусиралиева², К. Багитова^{2,3}, А.Т. Нюсупов²,
Е. Абайұлы²

¹Университет Пассау, Пассау, Германия;

²Казахский национальный университет им. аль-Фараби, Алматы, Казахстан;

³Атырауский университет им. Х. Досмухамедова, Атырау, Казахстан.

E-mail: *bolatbek.milana@gmail.com*

ФИШИНГОВЫЕ СООБЩЕНИЯ НА ВЕБ-РЕСУРСАХ И ИХ ОПРЕДЕЛЕНИЕ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Аннотация. В настоящее время в связи с увеличением использования веб-ресурсов увеличивается и количество видов киберпреступности. Среди них одним из наиболее распространенных видов преступлений является фишинг, фишинг – это совокупность способов обмана пользователей и получения их пароля, а также номера кредитной карты и другой конфиденциальной информации. Чаще всего злоумышленники позиционируют себя как сотрудников электронной почты или представителей известных организаций сотовых связей. Фишинг – один из самых простых, но самых опасных и эффективных способов кибератак. Фишинговые сообщения могут быть похожи на обычные легальные сообщения, когда злоумышленник может достичь своей цели, предоставляя пользователю Интернета искаженные веб-сайты, ссылки. В настоящее время антифишинговые подходы требуют от экспертов выявления признаков фишинговых сайтов и использования сторонних сервисов для идентификации фишинговых веб-ресурсов. В данной работе дается определение понятия фишинга, его видов. Дается обзор последних научных статей в данном направлении. Кроме того, авторы составили сборник фишинговых текстов для обучения и тестирования методов машинного обучения для определения фишинговых данных и использовали различные методы машинного обучения по данному набору данных.

Ключевые слова: веб-ресурс, кибербезопасность, фишинг, набор данных, классификация текста.

M. Bolatbek^{1*}, Sh. Musiralieva², K. Bagitova^{2,3}, A. Нюсупов², E. Abaiuly²

¹University of Passau, Passau, Germany;

²Al-Farabi Kazakh National University, Almaty, Kazakhstan;

³Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan.

E-mail: *bolatbek.milana@gmail.com*

PHISHING MESSAGES ON WEB RESOURCES AND THEIR DETECTION BY MACHINE LEARNING METHODS

Abstract. Currently, due to the increase in the use of web resources, the number of types of cybercrime is also increasing. Among them, one of the most common types of crimes is phishing, which is a set of ways to deceive users and obtain their password, as well as credit card numbers and other confidential information. Most often, attackers position themselves as an email employee or a representative of well-known cellular communications organizations. Phishing is one of the simplest, but most dangerous and effective methods of cyber attacks. Phishing messages can be similar to ordinary legal messages, when an attacker can achieve his goal by providing the Internet user with distorted websites, links. Currently, anti-phishing approaches require experts to identify signs of phishing web-sites and use third-party services to identify phishing web resources. This paper defines the concept of phishing, its types. An overview of the latest scientific articles in this direction is given. In addition, the authors compiled a dataset of phishing texts for training and testing machine learning methods to identify phishing data and used various machine learning methods for this dataset.

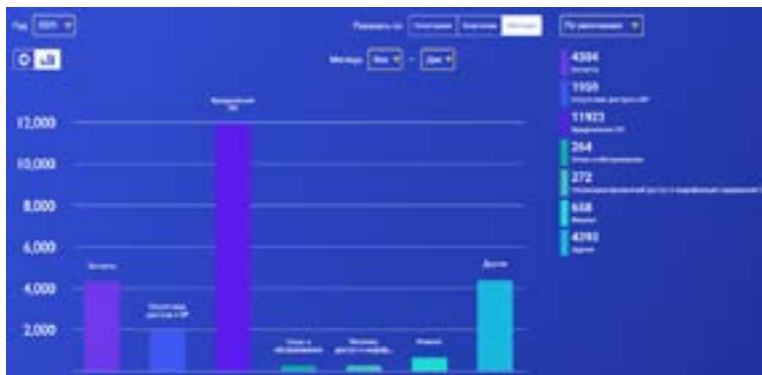
Key words: web-resource, cybersecurity, phishing, dataset, text classification.

Кіріспе. COVID-19 пандемиясы біздің өмірімізге үлкен өзгерістер алып келді, адамдардың жұмыс істеу, сауда жасау, қарым-қатынас орнату тәртібі және т.б. өзгеріске ұшырады. Вирустың таралуын шектеу мақсатында адамдарды өз-өзін оқшаулауға және әлеуметтік қашықтықты сақтауға мәжбүр болды, ал бұл өз кезегінде бүкіл әлемдегі Интернеттің, онлайн медианың қолданысын бірнеше есе арттырды.

Ақпараттық-коммуникациялық технологиялар қолданысының артуының жетістіктері өте көп, десек те аталған мүмкіндікті алаяқтар да өз пайдасына асыру үстінде. Ресми деректерге сәйкес әлемдегі, соның ішінде біздің елімізде де ақпараттық қауіпсіздік инциденттерінің саны күрт өскен. Солардың ішінде ең жиі кездесетін шабуылдардың бірі фишинг болып табылады. Қазіргі таңда фишинг – әлемдегі ең көп таралған киберқылмыс түрлерінің бірі, оның көмегімен аккаунттықжәне банктік деректер ұрланады.

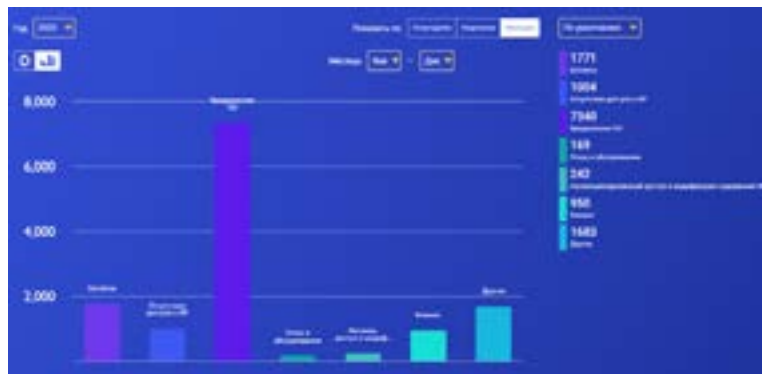
2021 жылдың қаңтар-желтоқсан аралығында елімізде тіркелген кибер-қауіпсіздік инциденттерінің саны 1-суретте келтірілген. Киберқауіпсіздік инциденттері ботнеттер, интернет ресурсқа қол жетімділіктің болмауы,

зиянды бағдарламалық жабдықтама, қызмет көрсетуден бас тарту, интернет ресурсқа рұқсатсыз қол жеткізу және оның мазмұнын модификациялау, фишинг және т.б. санаттарға жіктелген. Соның ішінде 658 фишинг шабуылы тіркелген.



1-сурет. 2021 жылы орын алған киберқауіпсіздік инциденттерінің статистикасы (KZ-CERT деректері негізінде)

KZ-CERT қызметінің деректері бойынша 2022 жылы алдыңғы жылғы көрсеткішпен салыстырғанда басқа киберқауіпсіздік инциденттенің саны азайса, фишингтік шабуылдардың саны күрт өскен, 2022 жылдың қаңтар айы мен осы уақытқа дейінгі аралықта 950 фишинг шабуылы анықталған, бұл бір жыл бұрынғыға қарағанда 1,4 есе көп (2-сурет).



2-сурет. 2022 жылға арналған ақпараттық қауіпсіздік инцидент түрлерінің динамикасы (KZ-CERT деректері негізінде)

Қазіргі таңда фишингтік веб-сайттар оларды анықтау дәлдігінің төмендігімен қауіпті болып табылады. Сол себепті фишингтік хабарламаларды уақытылы және дұрыс анықтауды киберқауіпсіздіктің маңызды саласы ретінде қарастыруға болады. Фишинг – бұл әлеуметтік инженерияның бір түрі, онда шабуылдаушы алаяқтық хабарлама жібереді, ол интернет пайдаланушысын

алдау арқылы шабуылдаушыға құпия ақпаратты ашуға мәжбүр етеді немесе зиянды бағдарламалық жасақтаманы пайдаланушының инфрақұрылымына, мысалы, ransomware сияқты орналастыруға арналады. 2020 жылдан бастап фишинг киберқылмыскерлер жасаған ең көп таралған шабуыл болып табылады, Федералды іздеу бюросы интернет-қылмыс туралы шағымдар орталығы компьютерлік қылмыстың кез-келген түріне қарағанда екі есе көп фишинг жағдайларын тіркейді.

Фишингтік оқиғалардың алдын алу немесе азайту әрекеттері заңнаманы білуді, пайдаланушыларды оқытуды, қоғамды ақпараттандыруды және техникалық қауіпсіздік шараларын қамтиды. Фишинг туралы хабардар болу үйде де, жұмыс орнында да маңызды болып табылады. Мысалы, 2017 жылдан бастап 2020 жылға дейін кәсіпорындар арасында фишингтік шабуылдар саны 72%-дан 86%-ға дейін өсті. Ғалымдар фишингтің бірнеше түрін ажыратады, олардың бірқатарына төменде қысқаша тоқталып өтеміз.

Электрондық пошта арқылы жүргізілетін фишинг

Фишингтік хабарламалардың көпшілігі электрондық пошта спама арқылы жеткізіледі және жекелендірілмеген немесе белгілі бір жеке тұлғаға немесе компанияға бағытталған – бұл "жаппай" фишинг деп аталады. Жаппай фишингтік хабарламаның мазмұны шабуылдаушының мақсатына байланысты әр түрлі болады – банктер мен қаржы қызметтері, электрондық пошта мен бұлтты қызмет провайдерлері және ағындық қызметтер жалпы мақсаттар болып табылады. Шабуылдаушылар деректерді жәбірленушіден ақшаны тікелей ұрлау үшін пайдалана алады, ұрланған ағындық қызмет шоттары әдетте қараңғы желі нарықтарындағы тұтынушыларға тікелей сатылады.

Тікелей (spear)фишинг Тікелей фишинг шабуылдаушы арнайы таңдалған фишингтік хабарламалар арқылы белгілі бір ұйымға немесе адамға тікелей бағытталған деп болжайды. Бұл электрондық поштаны заңды деп ойлау үшін белгілі бір адамға электрондық хаттар жіберуді білдіреді. Жаппай фишингтен айырмашылығы, тікелей фишинг арқылы шабуылдаушылар шабуылдың сәтті болу ықтималдығын арттыру үшін көп жағдайда өз мақсаттары туралы жеке ақпарат жинайды және пайдаланады.

Whaling және CEO қылмысы.

Фишингтің бұл түрі жоғары деңгейдегі басшыларға және басқа да жоғары деңгейдегі мақсаттарға бағытталған тікелей фишингтік шабуыл болып табылады. Бұл ұйымның басқа қызметкерлерін белгілі бір әрекетті орындауға мәжбүрлеу мақсатында, әдетте оффшорлық шотқа ақша аудару мақсатында жоғары деңгейдегі басшылардан жалған электрондық хаттар жасауды қамтиды.

Клон фишингі.

Клон фишингі – бұл фишинг шабуылының бір түрі, мұнда бірдей немесе клондалған электрондық поштаны құру үшін тіркеме немесе сілтеме бар заңды және бұрын жеткізілген электрондық поштаның мазмұны және алушылардың мекен-жайлары алынып, пайдаланылады.

Дыбыстық фишинг.

Дыбыстық фишинг немесе вишинг– фишингтік шабуылдарды жүргізу үшін телефонияны (көбінесе IP-телефония арқылы дауыстық байланысты) пайдалану. Зиянкестер көптеген телефон нөмірлерін тереді және көбінесе мәтінді сөйлеу синтезаторларын қолдана отырып жасалған автоматты жазбаларды ойнатады, олар банктік шоттармен немесе жәбірленушінің несие карталарымен алаяқтық әрекеттері туралы жалған мәлімдемелер жасайды. Жәбірленушіге шабуылдаушылар бақылайтын нөмірге қоңырау шалу ұсынылады, ол автоматты түрде алаяқтыққа "рұқсат беру" үшін құпия ақпаратты енгізуді немесе оларды ақпарат алу үшін әлеуметтік инженерияны қолдануға тырысатын адаммен байланыстыруды ұсынады.

SMS фишинг.

SMS фишингнемесе смишингэлектрондық пошта фишингіне ұқсас, тек шабуылдаушылар мәтіндік хабарламаларды жеткізу үшін ұялы телефон нөмірлерін пайдаланады. Фишингтік шабуылдар әдетте пайдаланушыны сілтемені басуға, телефон нөміріне қоңырау шалуға немесе SMS хабарлама арқылы шабуылдаушы ұсынған электрондық пошта мекен-жайына хабарласуға шақырады. Содан кейін жәбірленушіге жеке мәліметтерін беру ұсынылады; бұл көбінесе басқа веб-сайттар немесе қызметтер үшін тіркелгі деректері болып табылады.

Парақшаны басып алу.

Бұл пайдаланушыларды зиянды веб-сайтқа немесе сайтаралық сценарий арқылы эксплуатациялар жиынтығына бағыттау мақсатында заңды веб-парақшаларға нұқсан келтіруді қамтиды. Хакер веб-сайтты бұзып, бұзылған веб-серверге кіретін заңды пайдаланушыларды шабуылдау үшін *MPack* сияқты эксплуатациялар жиынтығын енгізе алады. Парақшаны басып алу көбінесе корпоративті нысандарға шабуыл жасаумен бірге қолданылады.

Күнтізбелік фишинг.

Мұндай жағдайда фишингтік сілтемелер күнтізбелік шақыртулар арқылы жеткізіледі. Күнтізбеден шақыртулар жіберіледі, олар көптеген күнтізбелерге автоматты түрде қосылады. Бұл шақыртулар көбінесе жауапты жиналысқа шақыру және басқа да кең таралған іс-шаралар сұранысы түрінде болуы мүмкін.

Әдебиеттерге шолу. Блокчейн технологиясының дамуы криптовалюта нарығының дамуына жол ашты және блокчейнді қылмыс әлемінде қолданылуы мүмкін платформалардың біріне айналдырды. Ең жиі кездесетін қылмыс түрлерінің бірі ретінде фишинг блокчейн платформалары мен қолданушыларының үлкен экономикалық шығынға ұшырауына алып келеді. Бұл жұмыста фишингтік аккаунттарды анықтауға арналған гибриді терең оқыту желілеріне негізделетін модель ұсынылады және оның тиімділігі эфириум ортасында дәлелденеді. Ұсынылатын модель транзакциялық жазбалардан алынған белгілер арасындағы өзара байланысты алу үшін нейрондық желілерге және мақсатты аккаунттың транзакцияларын талдауға негізделетін жаңа әдіс болып табылады. Эксперимент нәтижесінде ұсынылатын әдістің

фишингтік алаяқтық аккаунттарды жоғары дәлдікпен табатындығы анықталған (Wen және т.б., 2023).

Бұл жұмыста фишингтік электронды хаттарды анықтау және жіктеу үшін биогеографияға негізделетін терең оқыту арқылы оңтайландыру моделі ұсынылады. Ұсынылатын модельдің басты мақсаты – заңды және фишингтік электронды хаттарды ажырату. Электронды поштаны жіктеу үшін терең пайымдаулардың тиімді желісінің моделі пайдаланылады және оның тиімділігі гиперпараметрлерді баптау арқылы арттырылады. Ұсынылатын модель бірнеше бағалау параметрлері бойынша ағымдағы қолданыстағы әдістерден жоғары көрсеткіш көрсеткен (Dutta және т.б., 2023).

Бұл жұмыста фишингтік хабарламаларды анықтауға арналған үш семантикалық модель ұсынылады: мәліметтер қабатын біріктірудің көп масштабты моделі (MDF), нысандар қабатын біріктірудің көп масштабты моделі (MFF) және тереңдетіп біріктірудің көп масштабты моделі (MIF). Эксперименттер нәтижесінде үш модельдің де жоғары дәлдікпен жұмыс істейтіні анықталған. 6 айға созылған 3016 фишингтік веб-сайттарды белсенді бақылау нәтижесінде ұсынылатын модельдің нақты жағдайда анықтау тиімділігі дәлелденген (Lui және т.б., 2022).

Бұл мақалада фишинг мәселесін шешуге арналған жаңа әдіс ұсынылады. Алдымен берілген веб-парақшаның HTML және ашық мәтін бөлігіндегі терминдер жиілігі символдар деңгейінде TF-IDF әдісі арқылы анықталады. Сонымен қатар, ұсынылатын гиперсілтемелік белгілер веб-парақшаның контенті мен URL мекен-жайы арасындағы өзара байланысты білдіреді. Үлкен көлемдегі дайын мәтіндік корпустың болмауына байланысты зерттеушілер 60 252 веб-парақшадан тұратын мәліметтер жинағын құрастырған. Эмпирикалық нәтижелер ұсынылатын әрбір жеке белгінің фишингтік хабарламаларды анықтауда маңызды екендігін көрсетті, десе де барлық белгілерді біріктіру фишингтік сайттарды анықтау дәлдігін арттыратындығы анықталды (Aljofey және т.б., 2022).

Бұл жұмыста көп қабатты перцептрон (MLP)Hybrid Salp Swarm Jaya (HSSJAYA) көмегімен оқытылады және веб-сайттардың күмәнді, заңды немесе фишингтік екендігін анықтау үшін қолданылады. Гибридті алгоритмдер көмегімен оқытылған MLP-ді тестілеу үшін Salp Swarm (SSA) және Jaya алгоритмдері Cuckoo (CS), генетикалық алгоритм (GA) және Firefly алгоритмі арқылы оқытылған MLP-мен салыстырылған. Эксперименттер нәтижесінде HSSJAYA көмегімен оқытылған MLP-дің басқа алгоритмдермен салыстырғанда веб-сайттың фишингтік техникасын жақсырақ табатындығы анықталған (Erdemir және т.б., 2022).

Бұл жұмыстар авторлар сертификаттардан алынған жалпыға мәлім ақпарат, сонымен қатар фишингтік веб-сайттардың суреттері мен ресурстары негізінде 133 667 фишингтік веб-сайттардың өмірлік циклына талдау жүргізген. Зерттеушілер фишингтік веб-сайттарды олардың домендік атауларындағы шаблондар бойынша науқандарға топтастыру арқылы анықталған науқан-

дардың көп жағдайда бастапқы бірнеше күн ішінде, орта есеппен 12 күн ішінде жүргізілетінін анықтаған. Алынған нәтижелер фишингтік веб-сайттарды ерте анықтаудың артықшылықтары мен шектеулерін көрсетеді, атап айтқанда, веб-сайтты құру мен оны бұғаттау тізіміне енгізу арасындағы уақыт аралығына және мақалада талданған фишингтік науқандарда бірнеше аптаға дейін домендік атаулардағы үлгілердің өзгермейтініне қатысты қорытындылар алынған (Drugy және т.б., 2022).

Берілген жұмыста авторлар фишинг пен спам хабарламаларын анықтау үшін терең оқыту және табиғи тілді өңдеу әдістері қолданылады. Ұсынылған әдістің тиімділігі мәтіндік жинақ үшін LSTM моделі көмегімен жіктеу барысында орташа дәлдіктің 99%, MLP моделі үшін 94% болуымен дәлелденеді (Dewis және т.б., 2022).

Бұл жұмыста фишингтік веб-сайттарды анықтау және жіктеу үшін терең автокодтау (ODAE-WPDC) желісіне негізделетін оңтайлы модель ұсынылады. Ұсынылатын модель алдымен артық шудан тазарту үшін мәтіндерге алдын ала өңдеу процесін жүргізеді. Әрі қарай белгілер анықталады, олардың көмегімен жіктеу есебі орындалады, модельдің жұмыс өнімділігін арттыру үшін оңтайландыру алгоритмі қолданылады. Нәтижесінде ұсынылатын модельдің тиімділігі Kaggle репозиторийындағы фишингтік URL мәліметтер жинағы арқылы тексеріледі. Алынған нәтижелер ұсынылатын модельдің жұмыс өнімділігінің 99,28% екендігін көрсеткен (Alqahtani және т.б., 2022).

Бұл жұмыста зерттеушілер фишингтік веб-сайттарды анықтауда URL-дің қандай сипаттамаларының пайдалы болатындығын анықтау мақсатында белгілердің корреляциясы мен рекурсивті алынуын пайдаланады. Мақалада 48 және 87 нысаннан тұратын екі мәліметтер жинағы пайдаланылған. Бірінші сценарий қуатты болжаудың ұпайлық корреляциясы мен белгілердің рекурсивті алынуын, екінші сценарий максималды ақпараттың корреляция коэффициенті мен белгілердің рекурсивті алынуын, ал үшінші сценарий Спирмен корреляциясы мен белгілердің рекурсивті алынуын біріктіреді. Үш сценарий де белгілер жиыны кіші болған жағдайда да жоғары дәлдікті қамтамасыз етеді (Moedjahedy және т.б., 2022).

Әдетте фишингті анықтаудың бірінші тапсырмасы шынайы сайтқа ұқсайтын күмәнді парақшаның пайда болғандығын растау болып табылады. Содан кейін URL мекен-жайларды талдау механизмдері арқылы күмәнді веб-парақша туралы қорытынды жасауға болады. Шынайы веб-сайт пен оның күмәнді көшірмесін салыстыру логотип, тақырыптық жолақ, қаріп түсі және стилі арқылы жүргізіледі. Берілген жұмыста фишингті анықтау үшін логотипке негізделетін жаңа әдіс ұсынылады. Жаңа әдіс негізінде ұсынылатын механизм күмәнді логотипті барынша шынайы брендтің логотипіне жіктейді. Ұсынылатын әдістің тиімділігі Оңтүстік Азия аймағындағы танымал фишингтік брендтерге негізделетін жаңадан құрастырылған мәтіндік жинақ негізінде зерттеледі. Зерттеу нәтижелері пайдаланылған 5 машиналық оқыту әдістерінің ішінде ең жоғары тиімділік кездейсоқ орман ансамблі әдісін пайдалану барысында алынатындығын көрсеткен (Panda және т.б., 2022).

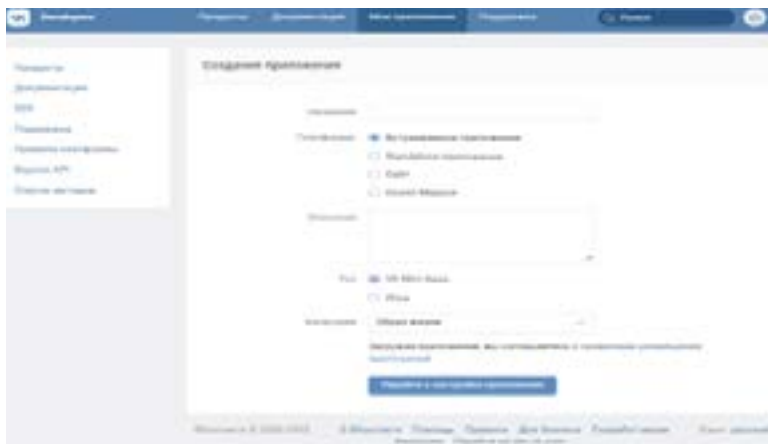
Берілген жұмыста зиянды мобильдік веб-парақшаларды анықтауға арналған тиімді әдіс ұсынылған. Авторлар ұсынатын әдіс белгілер векторын құру үшін берілген URL мекен-жайдан статикалық сипаттамалар мен сайттың танымалдық көрсеткіштерін алады. Содан кейін машиналық оқыту әдістері арқылы мәтіндер жинағындағы мәліметтерге жіктеу жүргізіледі. Кездейсоқ орман жіктеуіші басқа әдістермен салыстырғанда ең жоғарғы дәлдікті көрсеткен. Сонымен қатар, зерттеушілер қолданушыларға өз мобильді құрылғылары арқылы ұсынылатын жүйемен өзара әрекеттесуге мүмкіндік беретін қосымша құрастырған (Jain және т.б., 2022).

Материалдар мен әдістер. Веб-ресурстардағы фишингтік хабарламаларды машиналық оқыту әдістері арқылы анықтау үшін ең алдымен машиналық оқыту әдістерін оқыту және тестілеуге арналған мәтіндік корпус қажет болады. Берілген зерттеу жұмысында өзіндік фишингтік мәтіндер корпусы құрастырылып, пайдаланылды. Корпусты құрастыру барысында фишингтік бағыттағы және бейтарап мәтіндер заманауи бағдарламаларды қолдану арқылы жинақталды. Корпусты құрастыру барысында келесі әрекеттер орындалды: жинақталған корпустағы мәтіндерде фишингтік және бейтарап түрдегі мәтіндер үлесі шамалас болуын қадағалау; корпустағы мәліметтерге препроцессинг алгоритмдерін орындау; құрастырылған корпус бойынша машиналық оқыту алгоритмдерін оқыту; таңдалған әдіс бойынша жаңа ақпараттық ресурстарға талдау жүргізу.

Мәліметтерді жинау. Ең алдымен фишингтік ақпараттар жинақталады, әлеуметтік желілерден белгілі бір адамдардың назарын аудартуға арналған фишингтік хабарламалар жинақталды. Фишингтік мәтіндер адамдарды тегін ұтыстар мен жеңіл ақшаға шақыратын сайттар мен әлеуметтік желілердегі пікірлерден тұрады. Корпус жинақтау барысында «Youtube», «Вконтакте» және де басқа да веб-ресурстар қолданылды.

Веб-ресурстардағы мәліметтерді жинақтау үшін дайын парсер бағдарламалары қолданылды, олар бірнеше кезеңдер бойынша жұмыс атқарады және әрбір кілттік сөз бойынша жеке мәліметтер іздестіреді. Аталған бағдарламаға қандай іздеу жүйесі арқылы (Google, Yandex) мәліметтер жинақтау керек екенін, сонымен қатар қала, аймақ, қандайда бір мерзімді, іздеу тілін енгізе аламыз. Содан кейін белгілі бір алгоритм бойынша мәліметтер ізделіп, сүзгілеуден өткізіледі. Дайын есептерді қажетті форматта жүктеп алу мүмкіндігі бар.

Ақпарат жинау барысында «Youtube», «Вконтакте» секілді әлеуметтік желілермен басқа да веб-ресурстар таңдалды. Миллиондаған пайдаланушылар әлеуметтік желіде өздері туралы үлкен ақпаратпен бөліседі, статистикаға сәйкес, адамдар өздері туралы шындықты айтуды жөн көреді, сондықтан қандай да бір пайыздық қатынаста әлеуметтік желі парақшаларында көрсетілген ақпаратқа сенуге болады. Бұл ақпаратқа API арқылы қол жеткізуге болады. Vkontakte желісінде API алу арқылы өзімізге қажетті парақша түрін таңдау арқылы сол парақшадағы мәліметтер алынды (3-сурет).



3-сурет. «Вконтакте» әлеуметтік желісінен мәліметтер жинауға дайындық кезеңі

Препроцессинг кезеңі. Машиналық оқыту әдістерін оқытуға және тестілеуге арналған мәліметтер жинақталғаннан кейін ол мәліметтерді тазалау, токенизация, стоп сөздерді өшіру сияқты алдын ала өңдеу процестері орындалады.

Жоғарыда келтірілген функциялар арқылы барлық мәліметтер төменгі регистрге әкелінеді, содан кейін «шуль» таңбалар бос орындарға ауыстырылады. Осыдан кейін сілтемелер «URL», ал сандар «NUM» тіркестеріне ауыстырады.

Токенизация дегеніміз – бұл мәтіндерді токенге (жеке бөліктерге) бөлудің негізгі процесі.

Көбінесе күнделікті өмірде басқа сөздерге қарағанда қарағанда жиі кездесетін, сонымен қатар сөз тіркестерінің семантикалық көлеңкесіне ешқандай әсерін тигізбейтін сөздерді стоп-сөздер деп аталады. Бұндай сөздерді негізінен мәтіннен алып тастауға да болады, себебі бұл сөздер талдау барысында белгілі бір мағына бермейді. Құрастырылған корпустың фишингтік мәліметтері үшін құрастырылған сөздер бұлтын 4-суреттен көруге болады.



4-сурет. Корпустағы фишингтік мәліметтер классының сөздер бұлты

Құрастырылған корпустағы мәтіндер мысалы мен пайыздық үлесін келесі суреттерден көруге болады (5-6 суреттер):

label	message
0	greeting: Поздравляем, вы выиграли или Что скрывается за этой историей..
1	greeting: Поздравляем Вы стали победителем конкурса фото для ..
2	greeting: Поздравляем вы стали победителем автоконкурса на конкурсе..
3	greeting: Выиграли приз (Африка) в конкурсе..
4	greeting: Вы стали победителем Системы быстрых платежей..

5-сурет. Құрастырылған корпустағы фишингтік мәліметтерінің мысалы

label	message
403	neutral: Давид улетело ему зари тоном голосом тоном оверкило била. Ну что бы мне не видна.
404	neutral: Бурдакыйт паван в ту сторону отуда раздался ток но мило не расод и тавело водонул.
405	neutral: Зари покосило то зари подраженное его дикиме и урало в калитной старости адрити мравым.
406	neutral: Сидищия зари то тоном мравыному кад она пришло и подмало Харод что подол ольней атер и ольс зари в сторону.
407	neutral: А то бы в раскоридора в ступило бувако. Зол быто бы зари или бы зот заны кадры бувако покосило кубарем то даром.

6-сурет. Құрастырылған корпустағы бейтарап мәліметтерінің мысалы

Қазіргі таңда көптеген зерттеулерде фишингтік мәліметтерді анықтау үшін машиналық оқыту әдістеріне негізделетін алгоритмдер құрылған. Соған сәйкес берілген жұмыста бірнеше машиналық оқыту әдістері жиналған мәліметтер арқылы оқытылды және әр алгоритм бойынша жіктеу мәндері алынды. Мәтінді фишингтік және бейтарап класстарға жіктеу есебін шешу үшін Stochastic Gradient Descent, Random Forest, k-Nearest Neighbor, AdaBoost, Gaussian Naïve Bayes, Decision Tree сияқты алты түрлі машиналық оқыту әдістері пайдаланылды. Аталған әдістерді қолдану арқылы алынған жіктеу нәтижелері келесі бөлімде келтірілген.

Нәтижелер. Алдыңғы бөлімде айтылғандай, фишингтік және бейтарап мәтіндерді машиналық оқыту әдістері арқылы жіктеу барысында 6 түрлі әдіс пайдаланылды. Пайдаланылған әдістердің тиімділіктерін анықтау үшін Accuracy, Precision, Recall және F1-Score сияқты машиналық оқыту әдістерінің жұмыстарын бағалау параметрлері қолданылды.

Accuracy. Қарапайым жағдайда, мұндай метрика жіктеуіш дұрыс шешім қабылдаған құжаттардың үлесі бола алады.

$$Accuracy = P/N$$

мұндағы, P – жіктеуіш дұрыс шешім қабылдаған құжаттар саны, ал N – оқыту таңдамасының мөлшері.

Precision және Recall

Дәлдік пен толықтық (Precision және Recall) – бұл ақпарат алудың көптеген алгоритмдерін бағалауда қолданылатын өлшемдер. Кейде олар өздігінен қолданылады, кейде F-шара немесе R-Precision сияқты туынды метрикаларға негіз болады. Дәлдік пен толықтықтың мәні өте қарапайым болып табылады.

Жүйенің дәлдігі (Precision) дегеніміз – жүйенің берілген санатқа тағайындаған барлық құжаттарының шын мәнінде сол санатқа жататын құжаттар ішіндегі үлесі.

Жүйенің толықтығы (Recall) дегеніміз – жіктеуіш берілген санатқа жатады деп анықтаған құжаттардың тесттік жинақтағы сол санаттың барлық құжаттарына қатынасы.

F-өлшем

F өлшем – дәлдік пен толықтық арасындағы гармоникалық орташа мән. Егер дәлдік немесе толықтық нөлге ұмтылса, ол да нөлге ұмтылады.

$$F = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Бұл формула дәлдік пен толықтыққа бірдей салмақ береді, сондықтан F өлшемі дәлдік пен толықтықтың төмендеуімен бірдей болады.

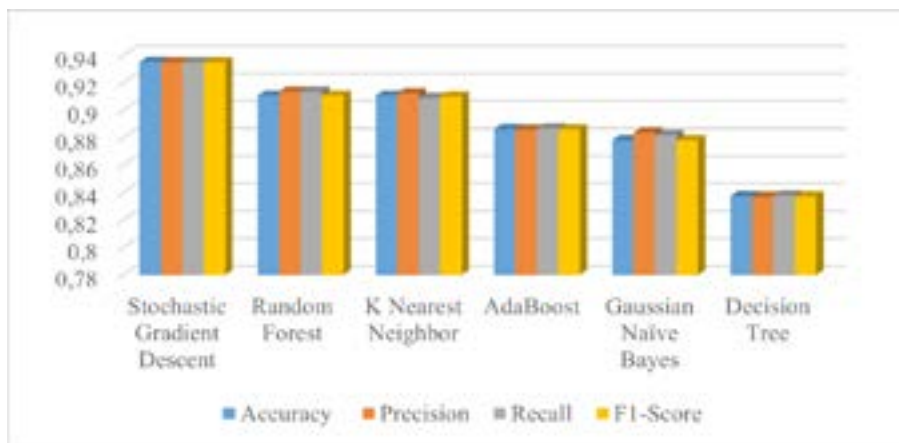
$$F = (\beta^2 + 1) \frac{\text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}}$$

Дәлдікке басымдық бергіңіз келсе мұндағы β , $0 < \beta < 1$ диапазонындағы мәндерді қабылдайды, ал $\beta > 1$ толықтыққа басымдық береді. $\beta = 1$ болған кезде формула алдыңғы формулаға келеді және теңдестірілген F өлшемін аламыз (оны F1 деп те атайды).

Stochastic Gradient Descent, Random Forest, k-Nearest Neighbor, AdaBoost, Gaussian Naïve Bayes, Decision Tree сияқты машиналық оқыту әдістері арқылы мәтінді фишингтік және бейтарап санаттарға жіктеу нәтижесі төмендегі сурет пен диаграммада көрсетілген (7-8 сурет).

model_name	accuracy_score	precision_score	recall_score	f1_score
Stochastic Gradient Descent	0.934959	0.934768	0.934768	0.934768
Random Forest	0.910009	0.912020	0.912020	0.910009
K Nearest Neighbor	0.910569	0.912166	0.909666	0.909674
AdaBoost	0.886179	0.882714	0.886737	0.885989
Gaussian Naïve Bayes	0.879049	0.883062	0.88183	0.878017
Decision Tree	0.837908	0.836906	0.837798	0.837129

7-сурет. Машиналық оқыту әдістері арқылы мәтінді жіктеу нәтижесі



8-сурет. Машиналық оқыту әдістерінің нәтижелерін салыстыру

Кесте мен суреттен көріп тұрғанымыздай, фишингтік хабарламаларды анықтауда пайдаланылған әдістердің ішінде барлық бағалау параметрлері бойынша ең жоғары көрсеткішті көрсеткен Stochastic Gradient Descent әдісі болып табылады, яғни веб-ресурстардағы фишингтік хабарламаларды анықтау есебін аталған әдісті қолдану арқылы шешуге болады деген қорытындыға келеміз.

Қорытынды. Берілген жұмыста қазіргі таңда ең жиі кездесетін және анықталуы жағынан бірқатар ерекшеліктері бар фишингтік хабарламаларды анықтау тапсырмасы қарастырылады. Зерттеу барысында машиналық оқыту әдістерін оқыту мен тестілеуге арналған өзіндік корпус құрастырылды. Фишингтік хабарламаларды анықтау есебі құрастырылған корпус арқылы машиналық оқыту әдістерін пайдалана отырып шешіледі және бірнеше әдіс ішінен тиімдісі таңдалады. Келешекте аталған мәтіндік корпусты кеңейтіп, фишингтік хабарламаларды терең оқыту әдістері арқылы анықтау есебін шешу қарастырылады.

Information about the authors:

M. Bolatbek – University of Passau, researcher, Passau, Germany. Senior Lecturer, Department of Information Systems, Al-Farabi Kazakh National University, PhD, Almaty, Kazakhstan, E-mail: bolatbek.milana@gmail.com, <https://orcid.org/0000-0002-2153-180X>;

Sh. Mussiraliyeva – Candidate of Physical and Mathematical Sciences, Head of the Department of Information Systems of the Al-Farabi Kazakh National University, Almaty, Kazakhstan. E-mail: mussiraliyevash@gmail.com, <https://orcid.org/0000-0001-5794-3649>;

K. Bagitova – Doctoral student of the Department of Information Systems of the Al-Farabi Kazakh National University, Almaty, Kazakhstan. Senior Lecturer, Department of Computer science, Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan, E-mail: kbbagitova@gmail.com, <https://orcid.org/0000-0003-1587-1995>;

A. Nyussupov – Doctoral student of the Department of Information Systems of the Al-Farabi Kazakh National University, Almaty, Kazakhstan. E-mail: moniumverse@gmail.com, <https://orcid.org/0000-0003-3254-0901>;

Y. Abaiuly – Doctoral student of the Department of Information Systems of the Al-Farabi Kazakh National University, Almaty, Kazakhstan. E-mail: erulan_97@mail.ru, <https://orcid.org/0000-0003-2248-3819>.

REFERENCES:

Aljofey A., Jiang Q., Rasool A. et al. An effective detection approach for phishing websites using URL and HTML features. *Sci Rep* 12, 8842 (2022). <https://doi.org/10.1038/s41598-022-10841-5>.

Alqahtani H., Alotaibi S.S., Alrayes F.S., Al-Turaiki I., Alissa K.A., Aziz A.S.A., Maray M., Al Duhayyim M. Evolutionary Algorithm with Deep Auto Encoder Network Based Website Phishing

Detection and Classification. *Applied Sciences*. 2022; 12(15):7441. <https://doi.org/10.3390/app12157441>.

Dewis M., Viana T. Phish Responder: A Hybrid Machine Learning Approach to Detect Phishing and Spam Emails. *Applied System Innovation*. 2022; 5(4):73. <https://doi.org/10.3390/asi5040073>.

Dong-Jie Liu, Guang-Gang Geng, Xin-Chang Zhang. Multi-scale semantic deep fusion models for phishing website detection, *Expert Systems with Applications*, Volume 209, 2022, <https://doi.org/10.1016/j.eswa.2022.118305>.

Erdemir E. i Altun A.A. (2022). Website Phishing Technique Classification Detection with HSSJAYA Based MLP Training. *Tehnički vjesnik*, 29 (5), 1696-1705. <https://doi.org/10.17559/TV-20211227132418>.

Jain A.K., Debnath N., Jain A.K. APuML: An Efficient Approach to Detect Mobile Phishing Webpages using Machine Learning. *Wirel Pers Commun*. 2022;125(4):3227-3248. doi: 10.1007/s11277-022-09707-w. Epub 2022 May 2. PMID: 35529800; PMCID: PMC9059682.

Kumar Dutta T. Meyyappan, Basit Qureshi, Majed Alsanea, Anas Waleed Abulfaraj, Manal M. Al Faraj, Abdul Rahaman Wahab Sait. Optimal Deep Belief Network Enabled Cybersecurity Phishing Email Classification Ashit, *Computer Systems Science & Engineering* DOI: 10.32604/csse.2023.028984.

Moedjahedy J., Setyanto A., Alarfaj F.K., Alreshoodi M. CCrFS: Combine Correlation Features Selection for Detecting Phishing Websites Using Machine Learning. *Future Internet*. 2022; 14(8):229. <https://doi.org/10.3390/fi14080229>.

Panda P., Mishra A.K., Puthal D. A Novel Logo Identification Technique for Logo-Based Phishing Detection in Cyber-Physical Systems. *Future Internet*. 2022; 14(8):241. <https://doi.org/10.3390/fi14080241>.

Tingke Wen, Yuanxing Xiao, Anqi Wang, Haizhou Wang. A novel hybrid feature fusion model for detecting phishing scam on Ethereum using deep neural network, *Expert Systems with Applications*, Volume 211, 2023, 118463, <https://doi.org/10.1016/j.eswa.2022.118463>.

Vincent Drury, Luisa Lux, and Ulrike Meyer. Dating Phish: An Analysis of the Life Cycles of Phishing Attacks and Campaigns. In *Proceedings of the 17th International Conference on Availability, Reliability and Security (ARES '22)*. Association for Computing Machinery, New York, NY, USA, Article 14, 1–11. 2022.

<https://doi.org/10.1145/3538969.3538997>.

https://www.researchgate.net/publication/346808880_Internet_use_and_well-being_during_the_COVID-19_outbreak_Examining_the_role_of_gender_age_motives_for_using_the_internet_and_relational_resources_in_an_Italian_adult_sample.

https://www.cert.gov.kz/press_club/infographics.

<https://en.wikipedia.org/wiki/Phishing>.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 30-42
<https://doi.org/10.32014/2022.2518-1726.154>

УДК 004.942

**М.А. Кантуреева^{1*}, А.Ш. Хасенов¹, Д.А. Тусупов¹, А.Б. Закирова¹,
А.З. Алимагамбетова²**

¹Евразийский национальный университет им. Л.Н. Гумилева,
Астана, Казахстан;

²Учреждение «Esil university», Астана, Казахстан.
E-mail: ma_khantore@mail.ru

FLOOR FIELD МОДЕЛЬ ДЛЯ ДИНАМИКИ ЭВАКУАЦИИ

Аннотация. В этой статье мы предложим метод построения статического полевого модели Floor Field для сложных помещений произвольной геометрии. Полевая модель Floor Field статического метода является важным компонентом модели и должно быть задано перед моделированием. Моделирование пешеходов – одна из самых захватывающих областей науки и техники о дорожном движении. Причина в том, что предварительное понимание характеристик пешеходного потока очень важно при проектировании или улучшении таких общественных мест, как залы ожидания на железнодорожных или автобусных станциях, супермаркеты, банкетные залы, конференц-залы, театры и кинотеатры. Динамические свойства пешеходных толп, включая различные явления самоорганизации, наблюдались и успешно воспроизводились различными физическими методами. Однако наблюдать за эвакуацией пешеходов гораздо сложнее, чем за обычным потоком пешеходов, из-за опасности и паники, вызванных инцидентами. Реальный эксперимент по эвакуации практически невозможен. Это побуждает исследователей изучать поведение при эвакуации с помощью различных подходов к моделированию. В этой статье мы улучшаем модель FF для моделирования эвакуации пешеходов в помещениях как с несколькими выходами, так и с внутренними препятствиями.

Ключевые слова: клеточный автомат, эвакуация, управления движением людей, множество клеток, оператор, клетка массива, динамическое поле, статическое поле.

Финансирование: Работа поддержана Комитетом науки МВОН РК, грант № AP08855497.

**М.А. Кантуреева^{1*}, А.Ш. Хасенов¹, Д.А. Тусупов¹, А.Б. Закирова¹,
А.З. Алимагамбетова²**

¹Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Ақпараттық технологиялар факультетінің «Ақпараттық жүйелер» кафедрасы, Астана, Қазақстан;

²«Esil university» мекемесінің Қолданбалы ғылымдар факультетінің Ақпараттық жүйелер және технологиялар кафедрасы, Астана, Қазақстан.
E-mail: ma_khantore@mail.ru

ЭВАКУАЦИЯ ДИНАМИКАСЫНА АРНАЛҒАН FLOOR FIELD МОДЕЛІ

Аннотация. Бұл мақалада күрделі геометрияның күрделі нысаналары үшін статикалық Floor Field үлгісін салу әдісі ұсынылады. Статикалық әдіс Floor Field модельдің маңызды құрамдас бөлігі болғандықтан, модельдеудің алдында анықталуы керек. Жаяу жүргіншілерді модельдеу жол қозғалысы ғылымы мен технологиясының ең қызықты салаларының бірі. Себебі, пойыз немесе автовокзалдардың күту залдары, супермаркеттер, банкет залдары, мәжіліс залдары, театрлар мен кинотеатрлар сияқты қоғамдық орындарды жобалау немесе жақсарту кезінде жаяу жүргіншілер қозғалысының ерекшеліктерін алдын ала түсіну өте маңызды. Жаяу адамдар тобының динамикалық қасиеттері, соның ішінде әртүрлі өзін-өзі ұйымдастыру құбылыстары әртүрлі физикалық әдістермен байқалды және сәтті шығарылды. Алайда, жаяу жүргіншілердің эвакуациясын бақылау оқиғалардан туындаған қауіп пен үрейге байланысты қалыпты жаяу жүргіншілер қозғалысына қарағанда әлдеқайда қиын. Шынайы эвакуация эксперименті іс жүзінде мүмкін емес. Бұл зерттеушілерді әртүрлі модельдеу тәсілдері арқылы эвакуациялау тәртібін зерттеуге шақырады. Бұл мақалада жаяу жүргіншілерді нысанада эвакуациялауды бірнеше шығумен және ішкі кедергілермен модельдеу үшін FF үлгісі жетілдірілді.

Түйін сөздер: клетка автоматы, скулшутинг, террористік шабуыл, эвакуация, адамдардың қозғалысын басқару, көптеген клеткалар, оператор, массивтің клеткасы, динамикалық өріс, статикалық өріс.

Қаржыландыру: Жұмысты ҚР ЖБҒМ Ғылым комитеті, №AP08855497 гранты қолдады.

**M. Kantureyeva^{1*}, A. Khassenov¹, D. Tussupov¹, A. Zakirova¹,
A. Alimagambetova²**

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

²Esil University, Esil University, Astana, Kazakhstan.

E-mail: *ma_khantore@mail.ru*

FLOOR FIELD MODEL FOR EVACUATION DYNAMICS

Abstract. In this article we will propose a method for constructing a static Floor Field model for complex rooms of arbitrary geometry. The Floor Field model of the static method is an important component of the model and must be specified before modeling. Pedestrian simulation is one of the most exciting areas of traffic science and technology. The reason is that a preliminary understanding of the characteristics of pedestrian traffic is very important when designing or improving public places such as waiting rooms at railway or bus stations, supermarkets, banquet halls, conference halls, theaters and cinemas. Dynamic properties of pedestrian crowds, including various phenomena of self-organization, have been observed and successfully reproduced by various physical methods. However, it is much more difficult to observe the evacuation of pedestrians than the usual flow of pedestrians, because of the danger and panic caused by the incidents. A real evacuation experiment is almost impossible. This encourages researchers to study evacuation behavior using various modeling approaches. In this article, we improve the FF model to simulate pedestrian evacuation in rooms with both multiple exits and internal obstacles.

Key words: cellular automaton, evacuation, human movement control, multiple cells, operator, array cell, dynamic field, static field.

Funding: *The work was supported by the Scientific Committee of the MHES of the RK, grant No. AP08855497.*

Введение. Клеточные автоматы (далее: КА) – это дискретные абстрактные вычислительные системы, которые оказались полезными как в качестве общих моделей сложности, так и в качестве более конкретных представлений нелинейной динамики в различных научных областях. Во-первых, КА (обычно) пространственно и временно дискретны: они состоят из конечного или исчисляемого набора однородных простых единиц, атомов или ячеек. В каждую единицу времени ячейки создают экземпляр одного из конечного набора состояний. Они развиваются параллельно с дискретными временными шагами, следуя функциям обновления состояния или правилам динамического перехода: обновление состояния ячейки выполняется с учетом состояний ячеек в ее локальной окрестности (следовательно, на расстоянии нет никаких действий). Во-вторых, КА абстрактны: они могут быть определены в чисто математических терминах, и физические структуры могут их реализовать.

В-третьих, КА – это вычислительные системы: они могут вычислять функции и решать алгоритмические задачи. Несмотря на то, что он функционирует иначе, чем традиционные устройства, подобные машине Тьюринга, КА с подходящими правилами может эмулировать универсальную машину Тьюринга и, следовательно, вычислять, учитывая тезис Тьюринга, что угодно вычислимое.

Клеточный автомат состоит из динамического правила, которое синхронно обновляет дискретную переменную, определенную на узлах (ячейках) d -мерной решетки. Одно и то же (локальное) правило перехода применяется ко всем ячейкам равномерно и одновременно. Правила КА могут быть выражены на языке, используемом для изучения динамических систем (Badii и др., 1997:49). Присвоение состояний всем ячейкам – это конфигурация. КА является обратимым (или реверсивным), если каждая конфигурация имеет ровно одного предшественника.

Могут быть определены различные типы окрестностей. В окрестности фон Неймана ячейки к северу, югу, востоку и западу от центральной ячейки определяются как соседи. Центральная ячейка имеет в общей сложности 5 соседей. Окрестность Мура также включает диагональные ячейки на северо-востоке, северо-западе, юго-востоке и юго-западе. С математической точки зрения КА представляют собой важный класс объектов, которые могут быть исследованы сами по себе и применяются для изучения природных явлений. Они ведут себя иначе, чем гладкие динамические системы, поскольку существование разрывов на сколь угодно малых интервалах препятствует какой-либо линеаризации динамики и, в свою очередь, простому исследованию свойств устойчивости траекторий. Тем не менее, некоторые инструменты, используемые для описания и классификации карт дискретного времени, могут быть расширены до КА; в частности, систематическое изучение повторяющихся конфигураций, поскольку они характеризуют асимптотическую эволюцию времени.

Необратимые КА представляют особый интерес для изучения сложных систем. При воздействии на начальное значение, состоящее из однородного фона из одинаковых символов, за исключением нескольких (возможно, одного), некоторые необратимые КА способны давать интересные предельные шаблоны, ни периодические, ни полностью неупорядоченные. Были предприняты попытки классифицировать многообразие КА как дискретные динамические системы. Вольфрам предложил различать четыре класса в соответствии с поведением КА; однако эта “феноменологическая” классификация страдает серьезными недостатками. В частности, Culik и Yu (Culik и др., 1988) показали, что нельзя решить, все ли конечные конфигурации данного КА становятся неподвижными и, следовательно, к какому классу он принадлежит. Были предложены и другие классификации. Например, Langton (Langton и др., 1990) предполагает, что правила КА могут быть параметризованы его параметром λ , который измеряет долю записей таблицы

правил без состояния покоя. Dubacq (Dubacq и др., 2001), а также Goldenfeld и Israeli (Goldenfeld и др., 2006) также предложили параметризацию правил КА, измеряемую их сложностью по Колмогорову.

КА можно рассматривать как модели распределенных динамических систем (Toffoli и др., 1984). Таким образом, они предстают как парадигма распределенных вычислений. Кроме того, можно отметить, что обратимый КА может выполнять обратимые, а именно сохраняющие информацию, вычислительные процессы.

Полевая КА-модель (Floor Field KA model). Широкое распространение получила полевая модель Floor Field (FF) (Song и др., 2005). В её основе лежит использование так называемых полей для моделирования движущей силы и взаимодействия с другими частицами.

Недавний прогресс в моделировании динамики пешеходов (Scheckenberg и др., 2001) замечателен, и многие ценные результаты получены с использованием различных моделей, таких как модель социальной силы (Helbing и др., 2000) и модель напольного поля (Burstedde и др., 2001) (Kirchner и др., 2002). Первая модель основана на системе связанных дифференциальных уравнений, которые должны быть решены, например, с использованием подхода молекулярной динамики, аналогичного изучению гранулированного вещества. Взаимодействие пешеходов моделируется с помощью сил отталкивания на большом расстоянии. В последней модели вводятся два вида полей пола, то есть статическое и динамическое, для перевода пространственного взаимодействия на большие расстояния в привлекательное локальное взаимодействие, но с памятью, аналогичное явлению хемотаксиса в биологии (Ben-Jacob и др., 1997). Интересно, что несмотря на то, что эти две модели используют разные правила для динамики пешеходов, они имеют много общих свойств, включая формирование полосы движения, колебания направления в узких местах (Burstedde и др., 2001) и так называемый эффект «быстрее-медленнее» (Helbing и др., 2000). Хотя это важные основы для моделирования пешеходов, еще многое предстоит сделать, чтобы применить модели к более практическим ситуациям, таким как эвакуация из здания со сложной геометрией.

В модели используются два поля: динамическое D и статическое S . Эти поля имеют такую же дискретную структуру, как и само пространство, по которому передвигаются частицы в КА моделях. Динамическое поле D соответствует виртуальному следу, который создается движением частиц по одной и той же ячейке и оказывает влияние на движение других. Оно имеет свою собственную динамику, а именно: рассеивание и забывание. Статическое поле S или поле расстояний не изменяется со временем, представляет своего рода карту местности, где каждая ячейка содержит обратное расстояние до ближайшего выхода с учетом всех недвижимых препятствий. То есть значение S увеличивается с приближением к выходу. Поле S не зависит от наличия или отсутствия частиц в рассматриваемой области.

В модели используются два поля: динамическое D и статическое S . Эти поля имеют такую же дискретную структуру, как и само пространство, по которому передвигаются частицы в КА моделях. Динамическое поле D соответствует виртуальному следу, который создается движением частиц по одной и той же ячейке и оказывает влияние на движение других. Оно имеет свою собственную динамику, а именно: рассеивание и забывание. Статическое поле S или поле расстояний не изменяется со временем, представляет своего рода карту местности, где каждая ячейка содержит обратное расстояние до ближайшего выхода с учетом всех недвижимых препятствий. То есть значение S увеличивается с приближением к выходу. Поле S не зависит от наличия или отсутствия частиц в рассматриваемой области. Переходные вероятности для всех частиц зависят от значений полей S и D в соседних клетках. Формула вычисления вероятностей устроена так, что наибольшая вероятность придается направлениям с наибольшими значениями поля.

При вычислении статического поля S (поля расстояний до выхода из любой точки пространства) могут использоваться несколько альтернативных методов. Для самых простых геометрий пространства (коридор, комната без дополнительных препятствий и поворотов) в (Song и др., 2005) поле S вычислялось посредством евклидова расстояния. Но для более сложных геометрий такой способ не подходит, так как необходимо учитывать препятствия и стены на пути к выходу. В (Ronchi и др., 2016) рассматривались различные способы вычисления поля расстояний (манхэттенское расстояние, расстояние Чебышева, применение алгоритма Дейкстры (Fukui и др., 1999)). Алгоритм Дейкстры более точен по сравнению с другими методами. Значения поля S влияют на динамику движения частиц в модели. В (Leng и др., 2014) сравнивались манхэттенское расстояние и алгоритм Дейкстры, при использовании алгоритма Дейкстры динамика движения частиц более правдоподобная. В FF моделях обычно используется обратное поле расстояний, т.е. значения поля S увеличиваются при приближении к выходу.

В FF моделях обычно используется синхронный режим работы КА: все частицы перемещаются на новые позиции одновременно. При таком режиме происходят конфликтные ситуации, когда две или более частицы претендуют на одну ячейку. Асинхронный режим исключает возникновение конфликтных ситуаций. Частицы для перемещения выбираются случайным образом (Yue и др., 2007).

Литературный обзор. В работах (Kretz и др., 2010) исследуются способы разрешения конфликтных ситуаций, возникающих при синхронном режиме работы КА. Идея состоит в следующем. С вероятностью $\mu \in [0,1]$ одной из частиц, которая выбирается случайно из всех кандидатов, разрешается переместиться в спорную ячейку, остальные остаются на месте. Это позволяет описывать эффект затруднения движения между частицами. Параметр μ можно интерпретировать как некоторый вид локального давления между частицами. Если значение μ близко к единице, тогда соседние частицы могут

сильно мешать друг другу достичь желаемых ячеек. μ называют фрикционным параметром. В простейшем случае при возникновении конфликтной ситуации случайным образом выбирается одна частица для перемещения, остальные остаются на прежних местах (Dijkstra и др., 1959) (Nishinari и др., 2010).

Формула для вычисления вероятностей переходов – самая вариативная деталь моделей движения людей. В классическую формулу не только вводятся новые элементы, но и изменяется сама формула. Но всегда используется компонента с полем S .

В (Kirchner и др., 2003) был введен эффект инерции. В (Yue и др., 2007) коэффициент k_s берётся с отрицательным знаком, чтобы частицы двигались в сторону понижения поля S . В (Varas и др., 2007) вероятность перехода в занятую ячейку не равна нулю, а уменьшается в два раза по сравнению с вероятностью в случае свободной ячейки, и приписывается текущей ячейке, что даёт возможность частице остаться на прежнем месте (в FF моделях запрещается переход в занятую ячейку и вероятность перехода в этом случае равна нулю). В (Varas и др., 2007) введён элемент паники, когда некоторой части частиц запрещают передвигаться.

В (Fruin и др., 1974) и др. присутствуют компоненты психологического отталкивания от стен и от других частиц. Влияние стен определяется их близостью к текущей частице (Kirchner и др., 2003), при этом задаётся максимальное расстояние, на котором стены могут влиять на выбор направления. Влияние частиц определяется посредством плотности частиц или с помощью специального поля, в котором отражена некоторая мера близости ячейки до других частиц (Hoogendoorn и др., 2003). В (Helbing и др., 2001) плотность определяется как количество частиц, находящихся в окрестности Мура от рассматриваемой. В (Berg и др., 1997) – количество частиц в направлении.

В (Kretz и др., 2010) предложена такая форма расчёта переходных вероятностей, где к тому же учитывается принадлежность частицы к определенной группе (например, семья).

В моделях, где время дискретно, важным является вопрос о переводе модельных временных шагов в естественные единицы измерения. В (Yue и др., 2007) предполагается, что один дискретный шаг осуществляется за 0,3 секунды, исходя из того, что средняя скорость человека составляет около 1,3 м/с и размер ячейки равен $0,4 \times 0,4 \text{ м}^2$. В (Kirchner и др., 2003) скорость 1,2 м/с и шаг равен 0,33 секунды. В (Varas и др., 2007) один дискретный шаг осуществляется за 0,4 секунды (размер ячейки - $0,4 \times 0,4 \text{ м}^2$, средняя скорость движения 1 м/с).

Материалы и методы. Модели движения людей микроскопического (индивидуального) типа должны воспроизводить как индивидуальные свойства движения людей при малых плотностях людского потока, так и массовые при больших плотностях. Учёт индивидуальных свойств движения людей при моделировании должен обеспечивать выполнение закономерностей

движения потока при различных плотностях в различных геометриях пространства (прямой путь, сужения, повороты и т.д.) для различных видов движения людских потоков (слияние, пересечение и т.д.).

Из всего многообразия свойств, присущих движению людей, выделяется следующий минимум:

- Цель движения – достижение цели следования;
- Движение людей – случайный процесс. Траектория одного и того же человека при прохождении одинакового маршрута каждый раз пусть немного, но отличается от предыдущих;
- Психологическое отталкивание: люди стараются избегать излишнего контакта друг с другом и не приближаться близко к стенам;

Люди выбирают кратчайший путь к цели следования. Если использование кратчайшего пути невозможно, то человек использует альтернативный путь. То есть при движении человек старается минимизировать длину пути или время пути.

Поведение человека в панических ситуациях. Сначала мы обсудим различные виды человеческого поведения в ситуациях паники. Люди в помещении пытаются эвакуироваться в случае пожара, используя свою собственную стратегию. Стратегии эвакуации до сих пор хорошо изучены, и мы приводим пример эксперимента по эвакуации, который был проведен в крупном супермаркете. В ходе эксперимента пожарная сигнализация и ложное задымление были установлены внезапно, и после того, как люди сбежали из здания, их опросили об их выборе путей эвакуации и т.д. Были собраны данные от более чем 300 человек.

Следующий список показывает статистику данных ответов:

1. Я сбежал в соответствии с указателями и инструкциями, а также трансляцией или руководством продавцов (46,7%).
2. Я выбрал направление, противоположное месту курения, чтобы как можно скорее спастись от огня (26,3%).
3. Я воспользовался дверью, потому что она была ближайшей (16,7%).
4. Я просто следил за другими людьми (3,0%).
5. Я избегал направления, в котором идут многие другие люди (3,0%).
6. Рядом с дверью было большое окно, и вы могли видеть снаружи. Это была самая “яркая” дверь, поэтому я воспользовался ею (2,3%).
7. Я выбрал дверь, к которой привык (1,7%).

0	$(i-1, j)$	0
$(i, j-1)$	(i, j)	$(i, j+1)$
0	$(i+1, j)$	0

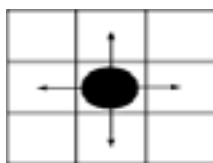


Рисунок 1. Целевые клетки для человека на следующем временном шаге. Для этой модели используется окрестность фон Неймана.

Мы видим, что были сделаны очень разные, иногда даже противоречивые, решения, указывающие на сложность проблемы эвакуации. Если мы предположим, что нет никаких знаков и указаний по радиопередачам, а также никакой информации о местоположении пожара, то, согласно анкетам, люди попытаются эвакуироваться, полагаясь как на свою память о маршруте до ближайшей двери, так и на поведение других людей. Это соревнование между коллективным и индивидуальным поведением имеет важное значение для моделирования явлений эвакуации.

Расширенная полевая модель. Здесь мы обобщим правила обновления расширенной модели поля метод для моделирования панического поведения людей, эвакуирующихся из помещения. Пространство дискретизировано на ячейки размером $40 \text{ см} = 40 \text{ см}$, которые могут быть либо пустыми, либо занятыми одним пешеходом (жесткое исключение). Каждый пешеход может перейти в одну из незанятых ячеек ближайшего соседа (i, j) (или оставаться в текущей ячейке) на каждом дискретном временном шаге $(t = t + 1)$ в соответствии с определенными вероятностями перехода $(P_{i,j})$.

В случае процессов эвакуации статическое поле метод S описывает кратчайшее расстояние до выходной двери. Напряженность поля $S_{i,j}$ устанавливается обратно пропорционально расстоянию от двери. Динамическое поле пола D — это виртуальный след, оставляемый пешеходами, подобный феромону при хемотаксисе (Ven и др., 1997). У него есть своя собственная динамика, а именно диффузия и распад, что приводит к расширению, разбавлению и, наконец, исчезновению следа. При $t = 0$ для всех узлов (i, j) решетки динамическое поле равно нулю, т. е. $D_{i,j} = 0$. Всякий раз, когда частица перескакивает из узла (i, j) в одну из соседних ячеек, D в исходной ячейке увеличивается на единицу.

Модель способна воспроизводить различные фундаментальные явления, такие как формирование полосы движения в коридоре, скопление и колебание в узком месте (Burstedde и др., 2001) (Kirchneг и др., 2002). Это незаменимое свойство для любой надежной модели динамики пешеходов, особенно для обсуждения вопросов безопасности.

Основные правила обновления. Правила обновления нашего КА имеют следующую структуру:

Поле динамического пола D модифицируется в соответствии с его правилами диффузии и затухания, контролируруемыми параметрами α и δ . На каждом временном шаге моделирования каждый отдельный бозон всего динамического поля D распадается с вероятностью δ и диффундирует с вероятностью α в одну из соседних ячеек.

Для каждого пешехода вероятности перехода $P_{i,j}$ для движения в незанятую соседнюю ячейку (i, j) определяются полями двух метода и собственной инерцией. Значения полей D (динамическое) и S (статическое) взвешиваются с двумя параметрами чувствительности k_D и k_S :

$$P_{i,j} = N \exp(k_D D_{ij}) \exp(k_S S_{ij}) P_1(i,j) P_w \quad (1)$$

с нормализацией N . Здесь p_i представляет эффект инерции (Burstedde и др., 2001), заданный $p_i(i,j) = \exp(k_i)$ для направления движения на предыдущем временном шаге, и $p_i(i,j) = 1$ для других ячеек, где k_i – параметр чувствительности. p_w представляет собой потенциал стенки. В (1) мы не учитываем ячейки препятствий (стены и т.д.), а также занятые ячейки.

Каждый пешеход случайным образом выбирает целевую ячейку на основе вероятностей перехода p_{ij} , определенных по формуле (1).

Всякий раз, когда два или более пешеходов пытаются переместиться в одну и ту же целевую ячейку, движение всех вовлеченных сторон отклоняется с вероятностью $\mu \in [0,1]$, т.е. все пешеходы остаются на своих площадках участники отклоняются с вероятностью $\mu \in [0,1]$, т.е. все пешеходы остаются на своих площадках (Kirchner и др., 2002). Это означает, что с вероятностью $1 - \mu$ одна из особей перемещается в нужную ячейку. Какому из них разрешено двигаться, решается с помощью вероятностного метода (Burstedde и др., 2001) (Kirchner и др., 2002).

Пешеходы, которым разрешено двигаться, совершают движение в целевую клетку, выбранную на шаге 3. D в исходной клетке (i,j) каждой движущейся частицы увеличивается на единицу: $D_{i,j} \rightarrow D_{i,j}+1$, т.е. –отрицательное целочисленное значение.

Вышеуказанные правила применяются ко всем пешеходам одновременно (параллельное обновление).

Вывод. Модели, основанные на теории клеточных автоматов, просты, понятны и способствуют более быстрому компьютерному вычислению (по сравнению с непрерывными моделями). Несмотря на активные исследования в данной области, остаются нерешёнными некоторые задачи.

Существующие КА-модели движения людей слабо формализованы (особенно, правила переходов). Описание модели проводится, как правило, в словесной форме, тогда как математическое описание позволяет однозначно интерпретировать модель и облегчает программную реализацию.

Поле расстояний в FF моделях используется в явном виде, что накладывает определённые ограничения на линейные размеры расчётной области.

Information about the authors:

Kantureyeva M. – Acting Associate Professor of the Department “Information Systems” of the Faculty of Information Technologies of the L.N. Gumilyov Eurasian National University, PhD. Astana, Kazakhstan, E-mail: ma_khantore@mail.ru. <https://orcid.org/0000-0001-5904-820X>;

Khasenov A. – Doctoral student of the Department “Information Systems” of the Faculty of Information Technologies of the L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: khssnv04@gmail.com, <https://orcid.org/0000-0002-6190-6776>;

Tussupov D. – Professor. Head of the Department “Information Systems” of the Faculty of Information Technologies of the L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: tussupov@mail.ru, <https://orcid.org/0000-0002-9179-0428>;

Zakirova A. – Candidate of Pedagogical Sciences, Associate Professor of the Department of Informatics of the L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: zakirova_ab@enu.kz, <http://orcid.org/0000-0001-8772-1414>;

A. Alimagambetova – Candidate of Physical and Mathematical Sciences, Senior Lecturer of Department of Computer Science, Esil University, Astana, Kazakhstan. E-mail: ainash_777@mail.ru, <https://orcid.org/0000-0002-9859-2029>.

ЛИТЕРАТУРА:

Кантуреева М.А., Тусупов Д.А., Мурзин Ф.А. Моделирование поведения толпы людей на базе клеточных автоматов // Вестник КазАТК. Серия технические науки. – 2019. – №2(109). – С. 168-175.

Кантуреева М.А. Моделирование движения толпы на основе клеточных автоматов в системе ANYLOGIC // Традиц. междунар. апрельская матем. конф. в честь Дня работников науки Республики Казахстан и Workshop «Problems of modelling processes in electrical contacts», посв. 80-летию юбилею С.Н. Харина. – Алматы, 2019. – С. 118.

Милинский А.И. Исследование процесса эвакуации зданий массового назначения: дис. ... канд. тех. наук. – М.: МИСИ, 1951. – 97 с.

A. Kirchner and A. Schadschneider, “Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics,” *Physica A*, vol.312, pp.260–276, 2002.

A. Kirchner, K. Nishinari, and A. Schadschneider, “Friction effects and clogging in a cellular automaton model for pedestrian dynamics,” *Phys. Rev. E* (in press) (e-print cond-mat/0209383).

A. Kirchner, H. Klüpfel, K. Nishinari, A. Schadschneider, and M. Schreckenberg, “Simulation of competitive egress behavior: Comparison with aircraft evacuation data,” *Physica A*, vol.324, p.691 (2003).

C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz, “Simulation of pedestrian dynamics using a twodimensional cellular automaton,” *Physica A*, vol.295, pp.507–525, 2001.

D. Helbing, I. Farkas, and T. Vicsek, “Simulating dynamical features of escape panic,” *Nature*, vol.407, pp.487–490, 2000.

D. Helbing, “Traffic and related self-driven many-particle systems,” *Rev. Mod. Phys.* vol.73, pp.1067–1141, 2001.

Dijkstra E.W. A note on two problems in connexion with graphs // *Numerische Mathematik*. – 1959. – Vol. 1. – P. 269-271.

E. Ben-Jacob, “From snowflake formation to growth of bacterial colonies II: Cooperative formation of complex colonial patterns,” *Contemp. Phys.* vol.38, pp.205–241, 1997.

Fukui M., Ishibashi Y. Self-organized phase transitions in cellular automaton models for pedestrians // *Journal of the Physical Society of Japan*. – 1999. – Vol. 68. – P. 2861-2863.

Kantureyeva M., Tussupov J., Murzin F., Uspanova A., Abisheva A. Application of cellular automata for modeling and review of methods of movement of a group of people // *Journal of Theoretical and Applied Information Technology*. – 2019. – Vol. 97, №15. – P. 4000-4010.

K. Abe, “Human Science of Panic” (in Japanese), Brain Pub. Co., Tokyo, 1986.

Kretz T., Bönisch C., Vortisch P. Comparison of various methods the calculation of the distance potential field // In book: *Pedestrian and Evacuation Dynamics* / ed. W.W.F. Klingsch, C. Rogsch, A. Schadschneider et al. – Berlin; Heidelberg: Springer-Verlag, 2010. – P. 335-346.

Leng B., Wang J., Zhao W., Xiong Z. An extended floor field model based on regular hexagonal cells for pedestrian simulation // *Physica A*. – 2014. — Vol. 402(0). – P. 119-133.

M. Scheckenberg and S.D. Sharma (Eds.), “Pedestrian and Evacuation Dynamics,” Springer-Verlag, Berlin, 2001.

M. de Berg, M. van Kreveld, M. Overmars and O. Schwarzkopf, “Computational geometry,” SpringerVerlag, Berlin, 1997.

Nishinari K., Suma Y., Yanagisawa D., Tomoeda A. et al. Nishi Toward smooth movement of crowds // In book: *Pedestrian and Evacuation Dynamics* / ed. W.W.F. Klingsch, C. Rogsch, A. Schadschneider, M. Schreckenberg. – Berlin; Heidelberg: Springer-Verlag, 2010. – P. 293-308.

Ronchi E., Kuligowski E.D. Assessing the verification and validation of building fire evacuation models // *Fire Technology*. – 2016. – Vol. 52. – P. 197-219.

S.P. Hoogendoorn, “Walker Behavior Modelling by Differential Games,” *Computational Physics of Transport and Interface dynamics*, Springer, 2003.

Song W., Yu Y., Fan W., Zhang H. A cellular automata evacuation model considering friction and repulsion // *Science in China Ser. E Engineering and Materials Science*. – 2005. – Vol. 48, №4. – P. 403-413.

V.L. Streeter, E.G. Wylie, and K.W. Bedford, “Fluid mechanics,” McGraw-Hill. 9ed., New York, 1985.

Varas A., Cornejo M.D., Mainemer D. et al. Valdivia Cellular automaton model for evacuation process with obstacles // *Physica A: Statistical Mechanics and its Applications*. – 2007. – Vol. 382(2). – P. 631-642.

Yue H., Hao H., Chen X., Shao C. Simulation of pedestrian flow on square lattice based on cellular automata model // *Physica A*. – 2007. – Vol. 384(2). – P. 567-588.

REFERENCES:

Kantureeva M.A., Tusupov D.A., Murzin F.A. Modeling the behavior of a crowd of people on the basis of cellular automata” // *Bulletin of KazATK. Series of technical sciences*. – 2019. – №2(109). – Pp. 168-175.

Kantureeva M.A. Modeling of crowd movement based on cellular automata in the ANYLOGIC system // *Traditional International. april math. conf. in honor of the Day of Science Workers of the Republic of Kazakhstan and Workshop “Problems of modeling processes in electrical contacts”*, dedicated to the 80th anniversary of C.N. Kharin. – Almaty, 2019. – p. 118.

Milinsky A.I. Investigation of the process of evacuation of mass buildings: dis. ... Candidate of Technical Sciences. – M.: MISI, 1951. – 97 p.

A. Kirchner and A. Schadschneider, “Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics,” *Physica A*, vol.312, pp.260–276, 2002.

A. Kirchner, K. Nishinari, and A. Schadschneider, “Friction effects and clogging in a cellular automaton model for pedestrian dynamics,” *Phys. Rev. E* (in press) (e-print cond-mat/0209383).

A. Kirchner, H. Klüpfel, K. Nishinari, A. Schadschneider, and M. Schreckenberg, “Simulation of competitive egress behavior: Comparison with aircraft evacuation data,” *Physica A*, vol.324, p.691 (2003).

C. Burstedde, K. Klauack, A. Schadschneider, and J. Zittartz, “Simulation of pedestrian dynamics using a twodimensional cellular automaton,” *Physica A*, vol.295, pp.507–525, 2001.

D. Helbing, I. Farkas, and T. Vicsek, “Simulating dynamical features of escape panic,” *Nature*, vol.407, pp.487– 490, 2000.

D. Helbing, “Traffic and related self-driven many-particle systems,” *Rev. Mod. Phys.* vol.73, pp.1067–1141, 2001.

Dijkstra E.W. A note on two problems in connexion with graphs // *Numerische Mathematik*. – 1959. – Vol. 1. – P. 269-271.

E. Ben-Jacob, “From snowflake formation to growth of bacterial colonies II: Cooperative formation of complex colonial patterns,” *Contemp. Phys.* vol.38, pp.205–241, 1997.

Fukui M., Ishibashi Y. Self-organized phase transitions in cellular automaton models for pedestrians // *Journal of the Physical Society of Japan*. – 1999. – Vol. 68. – P. 2861-2863.

Kantureyeva M., Tussupov J., Murzin F., Uspanova A., Abisheva A. Application of cellular automata for modeling and review of methods of movement of a group of people // *Journal of Theoretical and Applied Information Technology*. – 2019. – Vol. 97, №15. – P. 4000-4010.

K. Abe, “Human Science of Panic” (in Japanese), Brain Pub. Co., Tokyo, 1986.

Kretz T., Bönsch C., Vortisch P. Comparison of various methods the calculation of the distance potential field // In book: *Pedestrian and Evacuation Dynamics* / ed. W.W.F. Klingsch, C. Rogsch, A. Schadschneider et al. – Berlin; Heidelberg: Springer-Verlag, 2010. – P. 335-346.

Leng B., Wang J., Zhao W., Xiong Z. An extended floor field model based on regular hexagonal cells for pedestrian simulation // *Physica A*. – 2014. — Vol. 402(0). – P. 119-133.

M. Schreckenberg and S.D. Sharma (Eds.), “Pedestrian and Evacuation Dynamics,” Springer-Verlag, Berlin, 2001.

M. de Berg, M. van Kreveld, M. Overmars and O. Schwarzkopf, “Computational geometry,” Springer-Verlag, Berlin, 1997.

Nishinari K., Suma Y., Yanagisawa D., Tomoeda A. et al. Nishi Toward smooth movement of crowds // In book: *Pedestrian and Evacuation Dynamics* / ed. W.W.F. Klingsch, C. Rogsch, A. Schadschneider, M. Schreckenberg. – Berlin; Heidelberg: Springer-Verlag, 2010. – P. 293-308.

Ronchi E., Kuligowski E.D. Assessing the verification and validation of building fire evacuation models // *Fire Technology*. – 2016. – Vol. 52. – P. 197-219.

S.P. Hoogendoorn, “Walker Behavior Modelling by Differential Games,” *Computational Physics of Transport and Interface dynamics*, Springer, 2003.

Song W., Yu Y., Fan W., Zhang H. A cellular automata evacuation model considering friction and repulsion // *Science in China Ser. E Engineering and Materials Science*. – 2005. – Vol. 48, №4. – P. 403-413.

V.L. Streeter, E.G. Wylie, and K.W. Bedford, “Fluid mechanics,” McGraw-Hill. 9ed., New York, 1985.

Varas A., Cornejo M.D., Mainemer D. et al. Valdivia Cellular automaton model for evacuation process with obstacles // *Physica A: Statistical Mechanics and its Applications*. – 2007. – Vol. 382(2). – P. 631-642.

Yue H., Hao H., Chen X., Shao C. Simulation of pedestrian flow on square lattice based on cellular automata model // *Physica A*. – 2007. – Vol. 384(2). – P. 567-588.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 43-55

<https://doi.org/10.32014/2022.2518-1726.155>

FTAMP 50.41.29

А.Д. Кубегенова^{1*}, К.Т. Искаков², Е.С. Кубегенов^{1*}, О.И. Криворотько³

¹Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті,
Орал, Қазақстан;

²Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

³Новосібір мемлекеттік университеті», Ресей, Новосібір, Ресей.

E-mail: aigul-03@mail.ru

ДЕРЕКТЕРДІ ИНТЕЛЕКТУАЛДЫ ТАЛДАУ АРҚЫЛЫ ЭПИДЕМИОЛОГИЯЛЫҚ ЖАҒДАЙДЫ БАҚЫЛАУ ЖӘНЕ МОДЕЛЬДЕУ

Аннотация. Бұл зерттеу деректерді интеллектуалды талдау технологиясы, машиналық оқыту әдістері және кластерлік талдау негізінде есептерді сәйкестендіруді реттеу жүргізіледі, Қазақстанда әлеуметтік маңызы бар АИТВ ауруының таралуы бойынша математикалық модель үшін сандық шешу алгоритмдерін сипаттайды. АИТВ бойынша жағдайды модельдеудегі Data mining технологиясы ерекше өзекті болып табылады, өйткені оның негізінде Қазақстанда және еліміздің өңірлерінде сырқаттанушылықтың қысқа мерзімді болжамының карталары жасалады. Зерттеуде Қазақстандағы соңғы 10 жылдағына (2010-2020 жж.) талдау жасалды.

АИТВ-инфекциясының таралуы бойынша статистикалық деректер алынып қарастырылды. Ақпараттық технологиялар, соның ішінде Data mining технологиялары авторларға ауру кестесін сипаттауға, тәуекелдерді анықтауға, аурудың статистикалық болжамдарын тексеруге мүмкіндік берді. Зерттеудің негізгі бөлігінде сандық шешу алгоритмі және аймақтарды біртекті топтарға жіктеу арқылы АИТВ эпидемиологиясының математикалық моделін құру сияқты көрсеткіштер сипатталған.

АИТВ жұқтырғандарды өңдеу және олардың аймақтағы жағдайын талдау үшін Data Mining жіктеу әдістері қолданылды. Қазақстан халқының сырқаттанушылығын болжау Statistica қолданбалы бағдарламалар пакетінің көмегімен жүргізілді. Математикалық модельдеуге арналған сандық шешудің тиімді алгоритмі нақты деректердегі әзірлемелерді сынақтан өткізуге мүмкіндік береді.

Түйін сөздер: Statistica Advanced, Data Mining, кластерлеу, деректерді талдау, дендрограмма.

А.Д. Кубегенова^{1*}, К.Т. Искаков², Е.С. Кубегенов¹, О.И. Криворотько³

¹Западно-Казахстанский аграрно-технический университет имени Жангир хана, Уральск, Казахстан;

²Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан;

³Новосибирский государственный университет, Новосибирск, Россия.
E-mail: *aigul-03@mail.ru*

МОНИТОРИНГ И МОДЕЛИРОВАНИЕ ЭПИДЕМИОЛОГИЧЕСКОЙ СИТУАЦИИ С ПОМОЩЬЮ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Аннотация. В данном исследовании на основе технологии интеллектуального анализа данных, методов машинного обучения и кластерного анализа проводится корректировка идентификации задач, описываются алгоритмы численного решения для математической модели по распространенности социально значимого заболевания – ВИЧ-инфекции в Казахстане.

Технология Data mining в моделировании ситуации по ВИЧ является особенно актуальной, так как на ее основе разрабатываются карты краткосрочного прогноза заболеваемости в Казахстане и регионах страны. Исследование проводилось из взятых статистических данных по распространенности ВИЧ-инфекции в Казахстане за последние 10 лет (2010-2020г.).

Информационные технологии, в том числе технологии Data mining, позволили авторам описать графики заболеваний, выявить риски, проверить статистические прогнозы заболеваний.

В основной части исследования описаны такие показатели, как алгоритм численного решения и построение математической модели эпидемиологии ВИЧ путем классификации регионов по однородным группам. Для обработки ВИЧ-инфицированных и анализа их состояния в регионе были использованы методы классификации Data Mining. Прогноз заболеваемости населения Казахстана проводился с помощью пакета прикладных программ Statistica. Эффективный алгоритм численного решения для математического моделирования позволяет апробировать разработки на реальных данных.

Ключевые слова: StatisticaAdvanced, Data Mining, кластеризация, анализ данных, дендрограмма.

A.D. Kubegenova^{1*}, K.T. Iskakov², E.S. Kubegenov³, O.I. Krivorotko⁴

¹West Kazakhstan Agrarian and Technical University named after Zhangir Khan,
Uralsk, Kazakhstan;

²L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

³Novosibirsk State University, Novosibirsk, Russia.

E-mail: *aigul-03@mail.ru*

MONITORING AND MODELING OF THE EPIDEMIOLOGICAL SITUATION USING DATA MINING

Abstract. In this study, based on data mining technology, machine learning methods and cluster analysis, the identification of tasks is corrected, numerical solution algorithms for a mathematical model for the prevalence of socially significant HIV infection in Kazakhstan are described.

Data mining technology in modeling the HIV situation is particularly relevant, as it is used to develop maps of short-term prognosis of morbidity in Kazakhstan and the regions of the country. The study was conducted from statistically taken data on the prevalence of HIV infection in Kazakhstan over the past 10 years (2010-2020).

Information technologies, including Data mining technologies, allowed the authors to describe disease schedules, identify risks, check statistical forecasts of diseases.

The main part of the study describes such indicators as the numerical solution algorithm and the construction of a mathematical model of HIV epidemiology by classifying regions into homogeneous groups. Data Mining classification methods were used to process HIV-infected people and analyze their condition in the region. The prognosis of the morbidity of the population of Kazakhstan was carried out using the Statistica application software package. An effective numerical solution algorithm for mathematical modeling allows testing developments on real data.

Key words: StatisticaAdvanced, Data Mining, clustering, data analysis, dendrogram.

Қысқашы. Қазіргі уақытта Қазақстанда АҚТҚ жұқтырғандардың өсуіне байланысты қоғамдық денсаулық сақтауда жедел шешуді және көлемді деректерді талдауды талап ететін мәселелер жиі туындауда. Бұл проблема пандемия қаупінен басқа, жедел шараларды қажет ететін әлеуметтік, медициналық және экономикалық салдарлардың кең спектрін тудырады. Аурудың қауіптілік сипаты, ең алдымен, еліміздің жас ұрпақтың және еңбекке қабілетті азаматтарды зақымдануынан тұрады.

ЮНЭЙДС (2021) сарапшыларының деректері бойынша АИТВ-мен өмір сүретін адамдардың жалпы әлемдік саны 38,4 млн. астам адамды құрады [33,9 млн.-43,8 млн.], ЖҚТБ-мен байланысты аурулардан қайтыс болған адамдардың саны 650 000 құрады [510,000-860,000] адам (ЮНЕЙДС 2022).

Біріккен БҰҰ бағдарламасының мәліметтері АИТВ пандемиясының салыстырмалы тұрақтылығын көрсетсе де, ауру деңгейі әлі де жоғары. 2019-2021 жылдардағы коронавирустық инфекция пандемиясы бақылаусыз сипатқа ие инфекциялар қаупінің айқын мысалы болып табылады және барлық мемлекеттердің бұрын-соңды болмаған шараларымен еңсерілген міндеттерді бүкіл адамзатқа көрсетті. АИТВ-инфекциясының проблемасы таралудың басқа сипатына ие болса да, өте қауіпті болып қала береді, өйткені емдеудің жетістіктері ремиссия кезеңіндегі науқастардың түпкілікті қалпына келуіне әкелген жоқ. ЖИТС-тың алдын алу және оған қарсы күрес жөніндегі республикалық орталықтың деректері бойынша 30.09.2020 ж. өсу қорытындысымен АИТВ-инфекциясының 27 100 жағдайы тіркелді, оның ішінде ерлер-16344, әйелдер-10756, балалар — 146. Бұдан басқа, елімізде АИТВ оң статусы бар әйелдерден туылған 4464 бала тіркелген.

БҰҰ-ның ВИЧ/СПИД жөніндегі біріккен бағдарламасының (ЮНЭЙДС) жаңа стратегиясы 2030 жылға қарай әлемде ЖКТБ індетін тоқтатуға міндеттеме алды. Бұл Қазақстан Республикасының Денсаулық сақтау жүйесін дамытудың 2016-2030 жылдарға арналған "Денсаулық" мемлекеттік бағдарламасында көрініс тапты.

Бұл мәселенің өзектілігі Қазақстанда, әсіресе жұқтыру қаупі жоғары халық топтарында АИТВ-инфекциясы эпидемиясының сипатын зерттеу қажеттілігінен туындады. Эпидемиологиялық аурудың дамуын болдырмау үшін популяциядағы ауруды алдын-ала анықтауға мүмкіндік беретін терең талдау әдістері қолданылады. Қазақстан медицинасы оның қызметінің барлық салаларына статистикалық өңдеуді енгізу қажеттілігін түсінді. Алайда, статистикалық өңдеу құралдарының кеңінен енгізілуімен сапалы талдау ғана емес, сонымен қатар деректерді визуализациялау процестерін егжей-тегжейлі және терең зерттеу қажеттілігі туралы түсінік пайда болды. Статистикалық талдауға арналған бағдарламалар пакетін білу ғана емес, сонымен қатар оларды әр нақты жағдайға нақтылау қажет. Медициналық зерттеулер жүргізудегі зерттеушінің маңызды міндеті-деректерді статистикалық талдаудың нақты әдісін таңдау.

Медициналық ақпарат жүйесінің ауқымы мен күрделілігі күрт өсті және оны дамыту мен басқару жөніндегі қызметті бақылау қиын. Математикалық статистиканың дәстүрлі әдістері мен қарапайым әдістері саласында мәліметтер мен ақпараттың қарқынды өсуінен туындаған мәселелерді шешу қиын, бұл медициналық ақпараттық қызмет жүйесін басқаруға теріс әсер етеді. Сондықтан бағдарламалық жасақтаманы әзірлеу мен техникалық қызмет көрсетуді басқару үшін бағдарламалық жасақтама деректерін жинау өте маңызды болып келеді. Компьютерлік және ақпараттық технологиялардың, сондай-ақ сақтау технологиясының қарқынды дамуымен көптеген деректерді сақтауға болады. Деректерді іздеу технологиясы көптеген мәліметтерден потенциалды керек білімді іздеп және шығара алады. Деректер қоры технологиясы-бұл мәліметтер базасын басқаратын бағдарламалық жасақтама

туралы ғылым. Деректер базасындағы деректерді құрылымдау, жобалау және қолдану әдістерін зерттеу арқылы талданады (Xinui 2021). Ақпараттық технологияның қарқынды дамуымен мәліметтер базасының ауқымы, көлемі және тереңдігі кеңеюде, "бай деректер мен жаман ақпарат" деген ұғымға сәйкес келеді.

Деректерді іздеу дегеніміз-мәліметтер шаблонын іздеу процесі, яғни көптеген толық емес, анық емес, кездейсоқ мәліметтерден алынған мәліметтермен жұмыс жасау. Деректерді іздеу-бұл мәліметтер базасы мен жасанды интеллект саласындағы өте белсенді зерттеу саласы. Компьютерлік деректерді іздеу технологиясын әзірлеуге және қолдануға көп көңіл бөлу керек, өйткені деректерді іздеу технологияларын қолдана отырып, біз тұрақты дамуға ықпал ететін тиімді стратегияларды біріктіреміз (Pena-Ayala, 2014). Деректерді іздеу технологиясы тану параметрлерін және коэффициенттерді таңдауды егжей-тегжейлі талдайды, содан кейін деректерді іздеу моделі шығарылады.

Сандық векторлар жиынтығы ретінде, сипатталған әдістерге сәйкес кластерлерге топтастырылып Хопкинс статистикасының мәнімен есептеуге, кластерлеуді sklearn кітапханасының құралдарын қолдану арқылы жүзеге асыруға болады. Кластерлеудің екі түрлі K-Medium, АВТО-конфигурациясы бар тығыздыққа негізделген әдісі қолданылады. Салыстыру жағдайында кластерлік құрылым бір алгоритмнің әртүрлі параметрлерін өзгерту арқылы бағаланып (мысалы, k топтарының саны); алынған және дайындалған объектілерде модель (немесе бірнеше) құрылады және оның параметрлері түзетіліп, тестілеу және нәтижелерді талдау жүргізіледі. (Kubegenova және т.б., 2022)

Эпидемиологияның математикалық моделінің мысалы (АИТВ және туберкулездің коинфекциясы) математикалық модельдердің сәйкестендірілуіне арналған зерттеулерді көрсетеді. Модель параметрлерін анықтау міндеті квадраттық мақсатты функцияны азайтуға дейін азаяды (Romero және т.б., 2020).

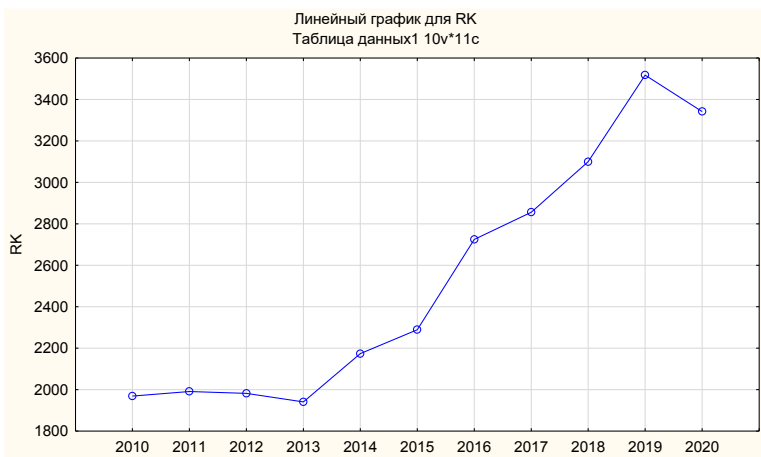
Сызықтық емес жүйелер қарастырылғандықтан, эпидемиологияның кері есептерін шешу екі түрлі болуы мүмкін, сондықтан кері есептердің сәйкестендірілуін талдау тәсілдері сипатталған. Бұл тәсілдер белгісіз параметрлердің қайсысын (немесе олардың комбинациясын) қолда бар Қосымша ақпарат бойынша бір мәнді және тұрақты қалпына келтіруге болатындығын анықтауға мүмкіндік береді (Zhenhua және т.б., 2020).

Зерттеу материалы мен әдістері. Зерттеу объектісі ретінде Қазақстан Республикасында АИТВ инфекциясымен сырқаттанушылықтың 10 жылдық кезеңінің (2010-2020 жж.) деректері таңдалды. Халықтың аурушандығы бойынша деректерді жіктеу Data Mining үлкен деректерін талдау арқылы жүргізілді. Деректерді талдау құралы ретінде Statistica қолданбалы бағдарламалар пакеті қолданылды: Statistica Base, StatisticaAdvanced, деректерді өндіруге арналған құралдар Data Mining, сондай-ақ SANN автоматтандырылған нейрондық желілер.

Кластерлеудің жаңа әдістері бір байланыс әдісіне негізделген графикалық формалардың көмегімен талдау жасауға мүмкіндік берді. Графикалық формаларды қолдану арқылы деректерді кластерлеу талдау уақытын қысқартуға, сондай-ақ сырқаттанушылықты болжау алгоритмін әзірлеуге мүмкіндік берді. Кластерлік талдауды деректерге қолданудың практикалық маңыздылығы мен өзектілігі күмән тудырмайды, өйткені қазіргі ақпараттық қоғамда деректер мен оларды талдау нәтижелері үлкен рөл атқарады, ал кластерлеу бұл деректерді жақсы түсінуге мүмкіндік береді. Деректерді талдау құралы ретінде Statistica қолданбалы бағдарламалар пакеті қолданылды: StatisticaBase, StatisticaAdvanced, Data Mining. Графикалық формаларды қолдану арқылы деректерді кластерлеу талдау уақытын қысқартуға, сондай-ақ сырқаттанушылықты болжау алгоритмін әзірлеуге мүмкіндік берді (Krivorotko және т.б., 2022).

Медициналық ақпаратты талдау үшін статистика әдістерін қолдану қазіргі уақытта Қазақстанда жеткілікті кең таралмаған, сондықтан біздің зерттеулеріміздің мақсаты Data Mining технологиясы бойынша эпидемиологиялық жағдайды талдау, болжау және алдын ала анықтау болып табылады. Эпидемиологиялық модельдің коэффициенттері халықтың ерекшеліктері мен аурудың дамуын сипаттайды. Экспериментталдық деректерден статистикалық деректерге квадраттың ауытқуы болып сипатталатыны, математикалық модельдегі параметрлерді сәйкестендірудің кері есебінің функциясының төмендеуі болып келеді. Статистикалық және оңтайландыру алгоритмдерінің жиынтығы 30% салыстырмалы дәлдікпен параметрлерді сәйкестендіруді көрсетеді. Нәтижелерді Денсаулық сақтау ұйымдары модельдеу деректерін тарихи деректермен салыстыру арқылы белгілі бір аймақтағы жұқпалы аурулардың эпидемиясын болжау үшін қолдана алады.

Нәтижелері және оларды талқылау. Эксперименттік мәліметтерді өңдеу компьютерде статистикалық пакеттерде жүргізілді.



Сурет 1. Қазақстан Республикасының 2010-2020 жылдарға арналған сызықтық графигі

Қазақстан Республикасының халқы бойынша (2010-2020 ж.) 10 жылдық кезеңнің жиынтық деректерін ескере отырып, АИТВ-инфекциясының (пациенттердің және тасымалдаушылардың саны) сырқаттанушылардың сызықтық графигі құрылды. (Сурет1)

Абсцисса осі бойынша АИТВ ауруын жұқтырған, зерттеу жылдары берілген, координаталар осі бойынша - АИТВ жұқтырғандардың абсолюттік саны келтірілген. (100 000 адам шаққанда) Бұл диаграммада аурудың 2010-2013 жылдар аралығында тұрақты тенденциясын көрсетеді. 2014 жылдан бастап аурудың нәтижесі екі есе көбейіп нашарлайды. 2019 жылы зерттеудің алғашқы жылдарымен салыстырғанда халықтың ауруы бірнеше есеге артып, шыңына жетті. Алайда, 2020 жылға қарай біз аурудың аздап төмендегенін байқаймыз. Бұл көрсеткіш бір жағынан жаңа короновирустық инфекция пандемиясы кезеңінде ақпарат жинау жүйесінің нашарлануымен, ал екінші жағынан оның өлім түріндегі салдарымен түсіндіріледі. Осылайша, Қазақстан Республикасында АИТВ-жұқпасымен сырқаттанушылықтың көпжылдық серпінін бағалау кезінде 2013-2019 жылдары жылдам көтерілу және 2019-2020 жылдар аралығында төмендеу анықталады. Тігінен берілген жолында 1,3% – ға дейін төмендеуі бұл тенденция абсолютті дегенді білдірмейді, өйткені онда ауытқу бар – нәтиже жақсарғанымен біртіндеп нашарлағанын байқаймыз. АИТВ-жұқтырғандардың сырқаттанушылық деңгейі бойынша сызықтық кестені талдау негізінде үш тапқа бөліп көрсетуге болады:

- 1) орташа деңгейден көтерілу жылдары (2010-2013);
- 2) жоғары деңгейден көтерілу жылдары (2013-2019);
- 3) құлдырау жылдары (2019-2020) аралығындағы жылдар (2014,2016,2018).

Variable	Descriptive Statistics (Исходные данные 2010-2020)							
	Valid N	% Valid obs.	Mean	Confidence 95.000%	Confidence 95.000%	Median	Mode	Frequency of Mode
Total revealed	11	100.000	46.4545	35.5122	56.9147	43.0000	43.0000	

Сурет 2. АИТВ жұқтырғандар нәтижелерінің сипаттама статистикасы бар электрондық кесте. (100 000 адамға шаққанда).

Бақыланатын айнымалының іріктемелі орташа мәні, төмендегі формула бойынша анықталады (1):

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

мұндағы n-үлгінің көлемі (айнымалы x бақылауларының нақты саны болып келеді). Медиан(Median) жоғарғы және төменгі жағынан екіге тең біркелкі бөлінген, реттелген мәндерден тұрады. Мода(Mode) -бұл мәліметтер жиынтығында жиі кездесетін мән.

Іріктемелі дисперсия айнымалының өзгергіштігін сипаттайды және формула бойынша есептеледі (2):

$$\overline{S}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (2)$$

Мұндағы \bar{x} - үлгінің орташа мәні. Дисперсия 0-ден шексіздікке дейін өзгереді. 0-дің төтенше мәні өзгергіштіктің болмауы – тұрақты айнымалылар.

Бастапқы деректер файлында Қазақстан Республикасының 16 өңірінде және 2 қаласында АИТВ жұқтырғандар туралы ақпарат бар. Осы кластерлік талдаудың мақсаты кластерлерге бөлу және тәуекел топтарын анықтау үшін тиісті кластерді анықтау болып табылады. Бұл мәселені шешу үшін кластерлік талдауды қолдану негізгі тиімді және жаппай қолданылатын әдістердің бірі болып саналады.

Жақындық шарасы ретінде Евклид арақашықтығын (Euclidean distances), ал кластерлерді біріктіру үшін – жалғыз байланыс әдісін (Single Linkage) немесе (жақын көршілерді) әдісті пайдалана отырып, кластерлік талдаудың иерархиялық рәсімінің көмегімен 16 өңірді жіктеуді жүргіземіз. Осы әдістердің көмегімен екі кластерді бір-бірімен байланыстыруға болады. Кез-келген екі кластер бір болғанда, олар бір-біріне жақындап және байланыс қашықтығымен ерекшеленеді.

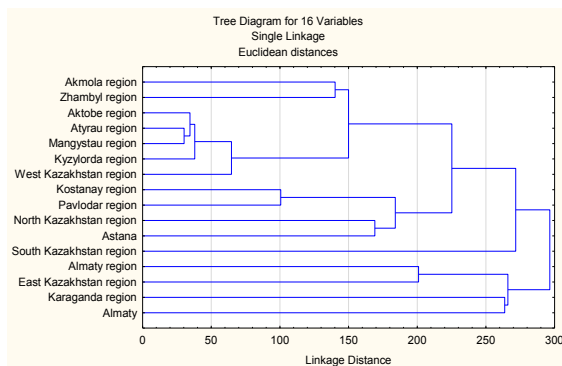
Сондықтан біріктірілген кластерлер кездейсоқ қалған бөліктерден бөлек элементтерге айналады. Бұл құбылыс нысандарды бір-бірімен байланыстырады және кластерлерді құрайды. Алынған кластерлер ұзын тізбектермен ұсынылған. Кластерлердің табиғи санын анықтау аймақтарды кластерлерге біріктіру арқылы жүргізілді. Өңірлерді кластерлерге біріктіру тәртібі иерархиялық ағашта берілген. Сурет 3.

$$\pi = \frac{a_i + a_j}{2b_{ij}} \quad (3)$$

Мұндағы a_i, a_j - сыныптардың орта кластық арақашықтығы;
 $i : j; b_{ij}$ - осы сыныптар арасындағы орташа кластық арақашықтықтар.
 Бағалауды дәстүрлі бөлу мынадай формула бойынша жүргізіледі:

$$S = \frac{1}{k} \sum_{i=1}^e \max \pi_{ij} \quad (4)$$

Жоғарыда сипатталған алгоритмнің көмегімен алынған бөліктер біреуіне тең немесе 1-ден аспайды. Сондықтан, барлық нысандар бір кластерге біріктіріліп, соңында біреуіне тең болады деп қорытынды жасауға болады.



Сурет 3. 2010-2020 жыл арлығындағы халықтың аурушандығы бойынша Қазақстан Республикасы өңірлерінің сыныптама дендограммасы.

Жалғыз байланыс (Одиночная связь, SingleLinkade) - әдісі ең ұғымды әдіс болып келеді және кең таралған атауы "Жақын көрші"(Ближайший сосед, NearestNeighbor). Алгоритмнің жұмысы ең жақын екі нысанды іздеумен ұсынылған, олардың комбинациясы бастапқы кластерді құрумен байланысты. Әрбір келесі объект осы жақын орналасқан кластерге қосылады. Объектілер жиынтығы бөлінген кластерлердің табиғи санын анықтау үшін иерархиялық кластерлеудің әр деңгейінде жиынтықты осы кластар санына бөлу жүргізілді. Кластерлердің әр жұбымен олардың бір-бірімен ішкі байланысы бағаланып, әр кластер үшін орташа кластерлік қашықтықты есептеу болып келеді. Байланысты бағалау ретінде орташа кластерішілік қашықтықтың кластераралық қашықтыққа қатынасы алынады. Дендограммада объектілер кластерлерге біріктірілетін қашықтықтар (шартты бірліктерде) көлденеңінен белгіленеді. Көлденең келген ось бақылауды, ал тігінен -біріктіру қашықтығын білдіреді (Сурет 3).

Алғашқы қадамдарда Қазақстан өңірлерінің кластерлері құрылуда: (Atyrauregion, Mangystauregion, Aktoberegion). Әрі қарай (WestKazakhstanregion, Kyzylordaregion) кластерлер пайда болды, бұл аймақтар арасындағы кластерлерді біріктіргенде үлкен болып келеді, алдыңғы қадамдарда қарағанда. (Pavlodarregion, Kostanayregion) келесі кластерлерге біріктіріледі (NorthKazakhstanregion, Astana). Бұдан әрі бір кластерге (Karaganda region, Almaty) және (Akmola region, East Kazakhstan region) және т. б. кластерлер біріктіріледі. Процесс барлық нысандарды бір кластерге біріктірумен аяқталады. Сонымен, дендограмма бойынша, бұл жағдайда үш кластерді бөлуге болады.(Кесте 1,2)

Кесте 1. Кластерлер арасындағы Евклид қашықтықтарының матрицасы.

	Кластер 1	Кластер 2	Кластер 3
Кластер1	0,0000	30464,68	101881,3
Кластер 2	174,5413	0,00	21707,3
Кластер 3	319,1884	147,33	0,0

Кесте 2. ҚР халқының аурушандығы туралы кластерлердің құрамы мен толықтырылуы 10 жылдық кезең бойынша (2010-2020 жж.).

Кластер №	Кластерді толықтыру, бірлік.	Кластер құрамы
1	3	Алматы облысы, Қарағанды облысы, Шығыс Қазақстан облысы.
2	5	Шығыс Қазақстан облысы, Астана қаласы, Солтүстік-Қазақстан облысы, Павлодар облысы, Қостанай облысы.
3	7	Батыс Қазақстан облысы, Қызылорда облысы, Манғыстау облысы, Атырау облысы, Ақтөбе облысы, Жамбыл облысы, Ақмола облысы.

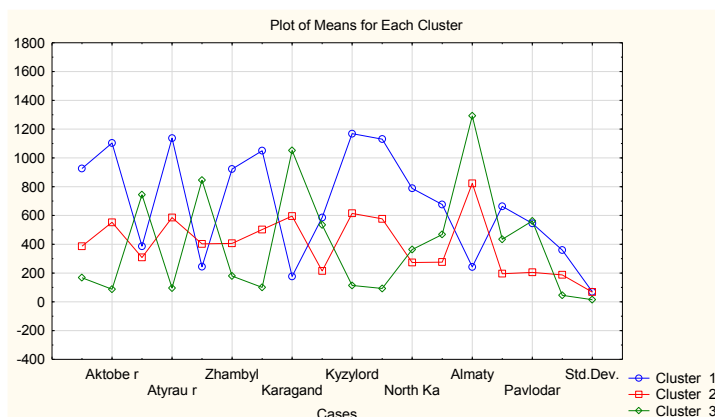
3 суретте көрсетілгендей 250 кесіндінің деңгейінде, 3 кластерге бөлінгенің байқаймыз. Алынған кластерлердің құрамы 2 кестеде анықталып көрсетілген. Алынған кластерлердің ерекшеліктерін талдап, аймақтардағы топтар (классы) бойынша АИТВ жұқтырғандар көрсеткіштерінің орташа мәнін салыстыра отырып, біз келесі нәтижелерге қол жеткіздік:

Бірінші кластер жалпы орта деңгейде кең таралған-АИТВ-инфекциясының жаңалықтары ересек тұрғындар мен халықтың осал топтары – есірткі тұтынушылар, сотталғандар арасында және АИТВ-инфекциясының жыныстық трансмиссиясының үлес салмағын алады;

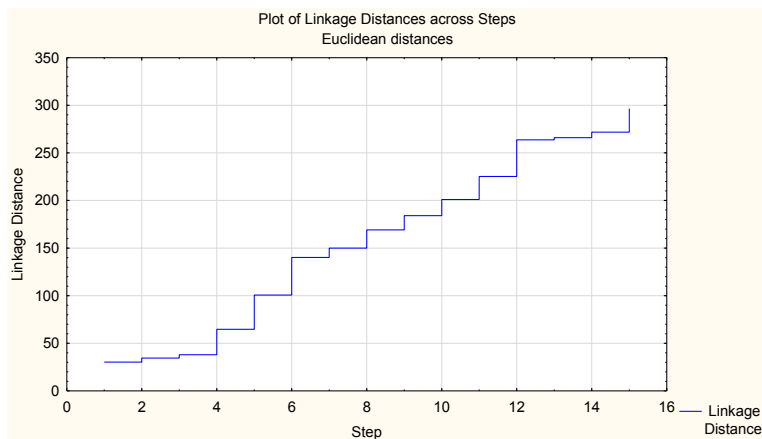
Екінші кластер үшін бірінші кластермен салыстырғанда АИТВ - жұқпанының төмен көрсеткішіне тән. Екінші кластерге тәуекелі жоғары, алайда алкогольмен, нашақорлықтан және басқа да әлеуметтік аурулардан зардап шегетін адамдар тобы кіреді;

Үшінші кластерде инъекциялық есірткіні тұтынушылар арасында АИТВ-инфекциясының таралуы мен берілу жолы, сондай-ақ ауру анадан құрсаққа жұқтыру жолдары көрсетілген.

3-суретте анықталғандай аймақтардың үш тобына қатысты айырмашылықтарды атап өтіп, 4 суретте әр кластер үшін орташа мәндерін көруге болады.



Сурет 4. Әр кластер үшін орташа мәндер графигі



Сурет 5. Қадамдармен біріктіру схемасының графигі.

Ағаш тәрізді (древовидная кластеризация) кластерлеу нәтижелері. Диаграммада ось бойынша көлденең қадамдар, тігінен – қашықтықты бейнелейді. Барлық нысандарды бір кластерге біріктіру үшін 5 суретте көрсетілгендей алгоритмге 16 қадам қажет болды.

Алынған жіктеме бірінші кластерге біріктірілген өңірлерде АИТВ жұқтырғандардың жоғары өсуі бар кластерлерді анықтады. Өңірлерді біртекті топтарға біріктіру және кері міндеттерді шешу жолымен алынған статистикалық болжамдау нәтижелері инъекциялық есірткіні тұтынушылар, сырқаттанушылықтың болжамдаушылары болып табылатынын көрсетті. Халықтың осы тобы АИТВ-инфекциясы індетінің өсуін ынталандыруды, жалғастырып отырғанын Data mining технологиясының көмегімен өңделгенін көрсетті. Жыныстық жолмен берілетін инфекциялар құрылымында маңызды рөл атқаратын коинфекциялар үлесінің артуы үлкен алаңдаушылық тудырады.

Қорытынды. Бұл мақалада Қазақстан Республикасында АИТВ инфекциясымен сырқаттанушылықтың 10 жылдық кезеңінің деректер алынып, халықтың аурушандығы бойынша Data Mining технологиясында талдау жүргізіліп, Statistica қолданбалы бағдарламалар пакетінің көмегімен талданды.

Зерттеудің нәтижелері төмендегі көрсеткіштерді береді:

1. Қазақстан Республикасында 10 жылдық кезеңде (2010-2020 жж.) АИТВ бойынша сырқаттанушылыққа жүргізілген талдау АИТВ - жұқтырғандардың күрт өсуін және сырқаттанушылықтың тұрақты үрдісін анықтауға мүмкіндік берді;

2. Кластерлік жіктеу алгоритмі объектілер арасындағы ішкі байланысты анықтады және АИТВ эпидемиологиясының математикалық моделінің дұрыстығын көрсетті;

3. Data mining көмегімен өңдеу Қазақстанда АИТВ инфекциясы бойынша сырқаттанушылықтың тұрақты өсуін көрсетті;

4. Статистикалық болжамның нәтижелері жоғары қауіп тобын анықтайтын және АИТВ індетінің өсуін ынталандыратын аурудың болжаушыларын анықтады.

5. Кластерлеу және алынған кластерлердің сипаты олардың жиынтығынан Қазақстанда АИТВ инфекциясымен күресте нақты антиретровирустық препараттарды және терапевтік схемаларды модельдеу, оңтайландыру және таңдау бойынша арнайы базаны қалыптастыруға мүмкіндік береді.

Information about the authors:

Kubegenova Aigul Dauletovna – master of technical sciences, senior lecturer of the West Kazakhstan agrarian and Technical University named after Zhangir Khan, Uralsk, Kazakhstan. aigul-03@mail.ru, <http://orcid.org/0000-0003-0156-7757>;

Iskakov Kazizat Takuadinovich – doctor of physical and mathematical sciences, professor of L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. kazizat@mail.ru, <http://orcid.org/0000-0002-3872-4026>;

Kubegenov Erlan Saginalievich – teacher of West Kazakhstan agrarian and Technical University, Uralsk, Kazakhstan. erlando78@mail.ru, <http://orcid.org/0000-0001-7424-264>;

Krivorotko Olga Igorevna Candidate of Physical and Mathematical Sciences Novosibirsk State University, Novosibirsk, Russia. krivorotko.olya@mail.ru, <http://orcid.org/0000-0003-0125-4988>.

REFERENCES:

Jianguo Liu & Sheng Zhou (2021) Application Research of Data Mining Technology in Personal Privacy Protection and Material Data Analysis, *Integrated Ferroelectrics*, 216:1, 29-42, DOI: 10.1080/10584587.2021.1911255.

Statistical collection. Health of the population of the Republic of Kazakhstan and the activities of healthcare organizations for 2020. [https://amanbol.KZ/news/vich-V-kazhstane-dannye/https://masa.media/ru/website/Bijalwan V, Kumar V, Kumari P et al \(2014\) KNN based machine learning approach for text and document mining. International Journal of Database Theory and Application. Vol.7, No.1 \(2014\), pp.61-70.http://dx.doi.org/10.14257/ijdt.2014.7.1.06](https://amanbol.KZ/news/vich-V-kazhstane-dannye/https://masa.media/ru/website/Bijalwan%20V,%20Kumar%20V,%20Kumari%20P%20et%20al%20(2014)%20KNN%20based%20machine%20learning%20approach%20for%20text%20and%20document%20mining.%20International%20Journal%20of%20Database%20Theory%20and%20Application.%20Vol.%207,%20No.%201%20(2014),%20pp.%2061-70.%20http://dx.doi.org/10.14257/ijdt.2014.7.1.06).

Kabanikhin S.I. (2019). Determination of the coefficient of nonlinear systems of ordinary differential equations using additional statistical information. *International Journal of Mathematics and Physics*, vol. 10, No. 1, pp. 36-42, June 2019. ISO 2409-5508. <https://doi.org/10.26577/ijmph-2019-i1-5>.

Kubegenova A.D., Zhakhiena A.G., Baigubanova S.K., Utyasheva G.S., Omarov A.N.(2022). Clustering and data mining on the example of hiv-infected people data (2022) *Journal of Theoretical and Applied Information Technology*, 100 (13), pp. 5010-5018.

Kubegenova A.D., Tukuadina A.I., Krivorotko O.I., Nurusheva zh.t. (2022) technology of Intelligent Analysis in predicting the epidemiological situation of HIV infection in the West Kazakhstan region. *National Academy of engineering RK habarshysy.2022zh*. No. 3 (85) pp. 28-42 <https://doi.org/10.47533/2020.1606-146X.174>Liu, L. (2020).

Development and Application of Computer Data Mining Technology. In: Abawajy J., Choo K.K., Islam R., Xu Z., Atiquzzaman M. (eds) *International Conference on Applications and Techniques in*

Cyber Intelligence ATCI 2019. ATCI 2019. Advances in Intelligent Systems and Computing, vol 1017. Springer, Cham. https://doi.org/10.1007/978-3-030-25128-4_121.

Liu M., Qu M. & Zhao B. (2017). Research and Citation Analysis of Data Mining Technology Based on Bayes Algorithm. *Mobile NetwAppl* 22, 418–426 (2017). <https://doi.org/10.1007/s11036-016-0797-2>.

Xinyi Wang. The Role of Data Mining Technology in Advertising Marketing. *J. Phys.: Conf. Ser.* 1744 042202, 2021. DOI: 10.1088/1742-6596/1744/4/042202.

Peña-Ayala, Alejandro.(2014). "Educational data mining: A survey and a data mining-based analysis of recent works." *Expert systems with applications* 41.4 (2014): 1432-1462.

Yang J, Li Y, Liu Q, et al. Brief introduction of medical database and data mining technology in big data era. *J. Evid Based Med.* 2020; 13: 57– 69. <https://doi.org/10.1111/jebm.12373>.

YUNEJDS Informacionnyj byulleten' 2022 <https://www.unaids.org/ru/resources/fact-sheet>.

He Wu, Gongjun Yan, and Li Da Xu. "Developing vehicular data cloud services in the IoT environment." *IEEE transactions on industrial informatics* 10.2 (2014): 1587-1595.

Zhenhua HUANG, Zhenyu WANG, Li JIANG, Rui ZHANG, Chang LEI, Xingwei LIU, Xiaohui XIE. (2020). Analysis of COVID-19 spread characteristics and infection numbers based on large-scale structured case data. *SCIENTIA SINICA Informationis*, Volume 50, Issue 12: 1882(2020) <https://doi.org/10.1360/SSI-2020-0029>.

Romero C., Ventura S. (2020). Educational Data Mining and Learning Analytics: An Updated Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. Vol. 10, no. 3, pp. 1-21, doi: <https://doi.org/10.1002/widm.1355>.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 56-67

<https://doi.org/10.32014/2022.2518-1726.156>

ӘОЖ 004.9

Г. Қалман¹, М.А. Самбетбаева^{1*}, Д.А. Ақтаева², А.С. Илюбаев^{2*}

¹Л.Н. Гумилев атындағы Еуразиялық ұлттық университеті,
Астана, Қазақстан;

²Абай Мырзахметов атындағы Көкшетау университеті, Көкшетау, Қазақстан.
E-mail: guljamal14@gmail.com

МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН АНАФОРАНЫ ШЕШУ МОДЕЛІ

Аннотация. Қазіргі кезде референцияны шешу табиғи тілдегі ең күрделі және әлі күнге дейін толық шешімін таппаған мәселелердің бірі, осыған қарамастан, көптеген лингвистердің және зерттеушілердің еңбектері талқыланып зерттеу жүргізіліп референцияны толық шешу модельдерін жобалауды талап етіп отыр.

Атап айтсақ, компьютерлік лингвистиканың алдында тұрған өзекті міндеттердің бірі мәтіндік құжаттарда әртүрлі нысандар туралы: адамдар, ұйымдар, оқиғалар, орындар және т.б., сондай-ақ олардың арасындағы байланыстар туралы мәліметтерді бөліп көрсету болып табылады. Әрбір Ақпараттық нысан белгілі бір ұғымға/пәндік аймақтың қатынасына сәйкес келетін белгілі бір құрылымға ие. Болашақта мәтінді өңдеу белгілі бір ресми тілде шектеулі және нақты сипатталған ақпараттық жүйе аясында жүзеге асырылады деп болжауға болады. Мәтінді автоматты өңдеудің маңызды элементтері анафориялық байланыстарды орнату және бір нысанның әртүрлі атауларын анықтау болғандықтан, белгілі бір контексте кез-келген объектіге, ұйымға, адамға және т.б., бірнеше рет айтылады.

Бұл зерттеу жұмысында машиналық оқыту әдістеріне негізделген қазақ тіліндегі анафоралық қатынасты шешу моделі ұсынылды. Қазақ тілінің негізгі тілдік ерекшеліктері ескеріле отырып жіктеу алгоритмдері пайдаланылды. Машиналық оқытуда тірек векторлық машиналарды(SVM) және Libsvm кітапханасы қолданылды. SVM-тің негізгі идеясы – шешім қабылдау жазықтығына гиперплант орнату, ең үлкен маргинализацияға қол жеткізу үшін оң даналар мен теріс даналарды оқшаулау.

Зерттеу нәтижелерін салыстыру мақсатында бірнеше эксперименттер жасалды. Сынақ деректері жинағы ретінде TengriNews.kz-тен жаңалықтар топтамасы және Ғ.Мұстафин әңгімелерінен үзінділер қолданылды. Эксперименттер нәтижелерін салыстыру мақсатында 2-кестеде көрсетілген нәтижелер алынды. Алынған нәтижелер қойылған міндеттерді шешуге ұсынылған модельдің тиімділігін растайды.

Түйін сөздер: анафора, антецедент, референциялық қатынас, машиналық оқыту, тірек векторлық машиналар (SVM).

Ғ. Қалман¹, М.А. Самбетбаева^{1*}, Д.А. Ақтаева², А.С. Илюбаев^{2*}

¹Евразийского национального университета им. Л.Н. Гумилева,
Астана, Казахстан;

²Кокшетауский университет имени Абая Мырзахметова,
Кокшетау, Казахстан.

E-mail: guljamal14@gmail.com

МОДЕЛЬ РАЗРЕШЕНИЯ АНАФОРЫ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Аннотация. В настоящее время решение референции является одной из самых сложных и до сих пор не полностью решенных проблем на естественном языке несмотря на то, что работы многих лингвистов и исследователей обсуждаются и требуют разработки моделей полного решения референции.

В частности, одной из актуальных задач, стоящих перед компьютерной лингвистикой, является выделение информации о различных объектах: людях, организациях, событиях, местах и т.д., а также о связях между ними в текстовых документах. Каждый информационный объект соответствует определенному отношению области понятия/дисциплины и имеет определенную структуру. Можно предположить, что в будущем обработка текстов будет осуществляться в рамках определенной информационной системы, ограниченной и четко описанной на определенном официальном языке. Важными элементами автоматической обработки текста являются установление анафорических связей и отождествление разных названий одного и того же объекта, например, какого-либо объекта, организации, лица и т.п., упоминаемых более одного раза в определенном контексте.

В данной исследовательской работе была предложена модель решения анафорических отношений в казахском языке на основе методов машинного обучения. Мы использовали алгоритмы классификации, учитывающие основные лингвистические особенности казахского языка. Машинное обучение использовало машины опорных векторов (SVM) и библиотеку Libsvm.

Основная идея SVM состоит в том, чтобы разместить гиперплоскость

на плоскости принятия решений, изолируя положительные экземпляры и отрицательные экземпляры для достижения максимальной маргинализации.

Было проведено несколько экспериментов для сравнения результатов исследований. В качестве набора тестовых данных мы использовали подборку новостей Tengrinews.kz и выдержки из рассказов Г. Мустафина. Для сравнения результатов экспериментов были получены результаты, представленные в таблице 2. Полученные результаты подтвердили эффективность предложенной модели для решения поставленных задач.

Ключевые слова: анафора, антецедент, референтные отношения, машинное обучение, машины опорных векторов (SVM).

G. Kalman¹, M.A. Sambetbayeva^{1*}, A.C. Ilyubayev², D.A. Aktaeva^{2*}

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

²Abay Myrzakmetov Kokshetau University, Kokshetau, Kazakhstan.

E-mail: guljamal14@gmail.com

ANAPHORA RESOLUTION MODEL BASED ON MACHINE LEARNING METHODS

Abstract. At present, solving referencing is one of the most difficult and still not completely solved problems in natural language, despite the fact that the works of many linguists and researchers are discussed and require the development of models for the complete solution of referencing.

In particular, one of the urgent tasks facing computer linguistics is to isolate in textual documents information about various objects: people, organizations, events, places, etc., as well as the relationships between them. Each information object corresponds to a certain concept/relationship of the subject area and has a certain structure. It can be assumed that in the future text processing will be carried out within a certain information system, limited and clearly described in a certain official language. Important elements of automatic text processing are the establishment of anaphoric relationships and the definition of different names for the same object, for example, for any object, organization, person, etc. In a particular context, repeated pronunciation.

This research paper proposes a model for solving anaphoric relations in the Kazakh language, based on machine learning methods. We used classification algorithms taking into account the main linguistic features of the Kazakh language. Machine learning used support vector machines (SVM) and Libsvm library.

The basic idea of SVM-establish a hyperplane on the decision plane, isolate positive and negative instances to achieve the greatest marginalization.

Several experiments were conducted in order to compare the results of the study. As a set of test data Tengrinews.kz-we used a selection of news and excerpts from

the stories of G. Mustafin. In order to compare the results of the experiments we obtained the results shown in Table 2. The results confirmed the effectiveness of the proposed model for solving the tasks.

Key words: anaphora, antecedent, reference relations, machine learning, support vector machines (SVMs).

Кіріспе. Табиғи тілдердің көпшілігі үшін мәтінді талдау кезінде ең үлкен қиындық морфологиялық талдаудан сәтті өнделетін жеке сөздерді талдау емес, сөздер арасындағы қарым-қатынастарды немесе, басқаша айтқанда, байланыстарды талдау болып табылады. Бұл мақалада референциялық қатынастар егжей-тегжейлі қарастырылады. Екі сөз бір нақты объектіні білдірсе, сілтеме бойынша байланысқан болып саналады. Бұл құбылыс әдетте анафора деп аталады. Атап айтқанда, есімдік анафора – сілтеулік байланысқа қатысушылардың бірі есімдік болатын анафора. Бұл құбылысты келесі сөйлемнің мысалында қарастырамыз.

Ерасыл мектептен келді, **ол** бірден ағасына көмектесуге цехқа кетті.

Мұндағы «ол» есімдігі «Ерасыл» деген жалқы есімді білдіреді. Есімдік сілтеп тұрған сөзді АНТЕЦЕДЕНТ деп атайды (бұл жағдайда антецедент «Ерасыл»), сілтеме жасау құралы анафора деп аталады (бұл жағдайда анафора «ол»). Есімдік анафорасының шешілуі оның төл сөзімен сәйкес келуі болып табылады.

Анафора мен корференцияны шешу міндеті 1960 жылдардан бастап белсенді зерттеліп келеді, дегенмен, шешілмеген мәселелері әлі де бар. Бұл мәселені шешудің негізгі тәсілдерін Р.Митковтың (Митков, 1999:281) және Elango P (Elango, 2006:8) зерттеулерімен Prokofyev R. (Prokofyev et al., 2015:458) жұмыстарынан көреміз. Ағылшын тіліне арналған анафораларды автоматты түрде шешу саласындағы зерттеулер 70 жылдары басталды. Виноградов, Уилкс, Хоббстың алғашқы әдістері мен жүйелері негізінен синтаксистік ақпаратқа негізделген ережелермен жұмыс істеді, сонымен қатар энциклопедиялық білім де кеңінен қолданылды. 80 жылдары бұрын бөлек қолданылған әртүрлі белгілерді біріктіру тенденциясы пайда болды. Э. Рич пен С. Луперфойдың еңбектерінде Дж. Карбонелла, Р. Митков гендер мен сандарды, синтаксистік және семантикалық қатынастарды үйлестіретін алгоритмдерді сипаттады.

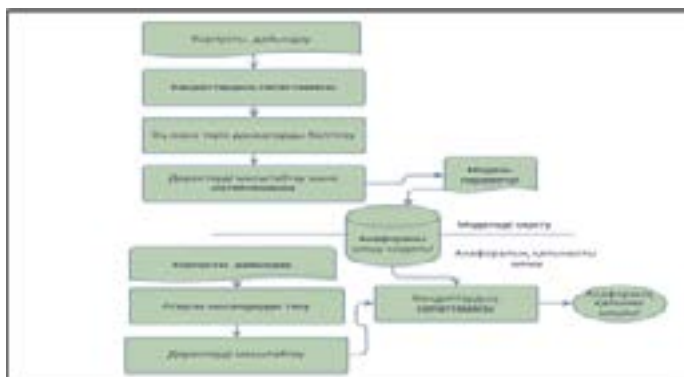
Орыс тіліне арналған анафораның шешу эксперименталды түрде аз зерттелген. Кибрик А. еңбектерінде (Кибрик, 2013: 118) орыс тіліндегі анафора құбылысының теориялық аспектілерін талқылайды және анафораның табиғатын көрсететін тілдік белгілер қатарын сипаттайды. Толпегин П (Толпегин, 2006: 88) еңбектерінде машиналық оқыту әдістерін қолдана отырып, орыс мәтіндеріндегі есімдік анафораны шешудің статистикалық моделін құру алгоритмін ұсынады. Абрамов В еңбектерінде (Абрамов, 2011:127) әлеуметтік-саяси мәтіндердің үйлесімділік ережелерін талдау үшін пайдаланатын әртүрлі сөйлемдер мен жағдайларда анафориялық қатынастарды анықтау принциптерін егжей-тегжейлі сипаттайды.

Қазіргі тәсілдер аннотацияланған корпустарды қолдана отырып, автоматты оқытуға негізделген. Олар дәстүрлі лингвистикалық әдістерді статистикалық әдістермен біріктіреді және морфологиялық, синтаксистік, семантикалық және тезаурус жиынтығы сияқты әр түрлі оқыту түрлерін қолданады. Бұл жұмыстың мақсаты қазақ тіліндегі анафоралық қатынасты тірек векторлық машиналар көмегімен шешу жолдарын зерттеу, қазақ тілінің тілдік ерекшеліктерін жіктеу, үлкен көлемдегі деректерді масштабтау мен деректерді оңтайландыру жолдарын ұсыну болып табылады.

Зерттеу әдістемесі мен материалдары. Жұмыс мақсатына жету үшін келесі зерттеу есептері қойылып оларды шешу жолдары қарастырылады:

Үлкен деректерді масштабтау, анафораның оң және теріс мәндерін алу, жіктеу алгоритмін құру, анафоралық шешу моделін жасау.

Машиналық оқытуға негізделген анафоралық ажыратымдылықтың жіктелу алгоритмі негізінен келесі төрт бөлікті қамтиды: корпусты алдын ала өңдеу, функция векторының құрылысы, оқыту классификациясының моделі және белгісіз қатынастар классификациясы. Қолдау векторлық машиналарының машиналық оқыту алгоритмінің сипаттамаларына сәйкес біз екі қадамды арттырамыз: оқыту деректерінің масштабтау процесі және оқыту параметрлерін оңтайландыру, 1-суретте көрсетілген жалпы процесс.



Сурет 1. Анафоралық қатынастарды шешу процесі

Тірек вектор моделі (svm) - тірек векторлық машиналары (Rong-en et al., 2005:1889) максималды маржа классификаторы болып табылады. $y \in \{1, -1\}$ i белгілері бар X_i $i = 1, 2, \dots$ даналарының жиынын ескере отырып, SVM-ді оқытудағы негізгі міндет келесі квадраттық оңтайландыру есебін шешу болып табылады:

$$\min_a f(a) = \frac{1}{2} a^T Q a - e^T a \quad (1)$$

$$0 \leq a_i \leq C, i = 1, \dots, i, y^T a = 0$$

Мұндағы e -барлық бірліктердің векторы, C - барлық айнымалылардың жоғарғы шегі;

Q-симметриялық матрица L-ге $Q_Y = y_i y_j k(X_i X_j)$ және $k(X_i X_j)$ ядро функциясы болып табылады.

Тірек векторлық машиналар негізінен жіктеу алгоритмінің идеясын қолданады, ол QR(квадраттық бағдарламалау мәселесі) (Zhang et all., 2003:32) мәселесін бірнеше кіші ішкі тапсырмаларға ыдыратады, бұл ішкі есептерді итеративті түрде шешуге және QR туралы оңтайлы шешім береді. Осы жіктеу алгоритмінің идеясының арқасында жіктеу стратегиясы, конвергенция шарттары және конвергенция сияқты басқа да мәселелерді шешуде, уақытты үнемдеу және жылдамдықты арттыру үшін жіктеу алгоритм орындалады.

SVMs-тің негізгі идеясы-шешім қабылдау жазықтығына гиперплант орнату, ең үлкен маргинализацияға қол жеткізу үшін оң даналар мен теріс даналарды окшаулау.

Табиғи тілді өңдеудің көптеген мәселелерін жіктелген мәселе ретінде қарастыруға болады; Жұмыстың негізгі мәселесін шешу идеясы оны классификацияға айналдыру болып табылады. Негізгі анафоралық қатынасының оң даналары мен теріс даналарына үміткерді болжаған кезде, SVM алгоритмі оң даналарды және теріс даналарды дұрыс ажырататын жіктеу сызығын табу және жіктелген интервалды максимум ету, бұл оңтайлы жіктеу сызығы деп аталады. Көп өлшемді кеңістіктің оқу деректерінде классификация қатесін минималды етіп жасай алатын категория бетін табу керек, оны оптималды гипер жазықтық деп атайды.

Векторлық Функцияның құрылысы – бұл мақаланың негізгі идеясы – негізгі анафоралық қатынасты түрлендіру классификацияға қатысты, содан кейін машиналық оқытуды пайдалану. Табиғи тілді өңдеуде - бұл үлкен көлемдегі «шектеулерді» қамтамасыз ету арқылы ғана сапалы талдау мен мәселені оңтайлы шешуге мүмкіндік бар. Бұл жұмыста келесі категорияларды қамтитын белгілерді қарастырады: сөздік белгілер, сөздер арасындағы қашықтық, жүйелілік белгілер және т.б.

Сөздік белгілер – екі үміткердің жолдарының сәйкестік дәрежесін білдіреді; жалпы алғанда, кандидаттардың жолдарының сәйкестік дәрежесі неғұрлым жоғары болса, кандидат сызықтарының анафоралық байланысы соғұрлым жоғары болады. Ол әдетте сөздіктің нобайы туралы ақпаратты, сөздіктің ішкі жолы туралы қысқартылған ақпаратты шығарып, сөздіктің сәйкестігі туралы толық ақпаратты қамтиды. Мысалы, "Елбасы" – бұл "президент", яғни бұл екі сөз бір бірімен анафоралық қатынастарды қалыптастырады.

Сөздер арасындағы қашықтық – екі үміткердің қашықтығын қарастырады; екі үміткердің арасындағы қашықтық неғұрлым жақын болса, кандидаттар арасындағы ықтималдық соғұрлым жоғары болады. Тұтастай алғанда, екі кандидат үштен көп сөйлеммен бөлінген, бұл негізгі анафоралық қатынас мүмкіндігі өте аз болады.

Жүйелілік белгілер – көптелу, гендерлік категория, семантикалық категорияға және т.б. бөлуге болады. «Біз» және «студенттер» сияқты барлық көпше түрдегі зат есімдер, олар жеке адамға сілтеме жасай алмайды.

Қазақ тілінде басқа тілдер сияқты гендерлік категорияға бөлінбейді, сөйлемдегі сөздің жалпы мағынасына және адам аттарының ерекшеліктеріне қарай біз сілтеменің ер немесе әйел не басқада объектілерге қатысты айтылғанын біле аламыз, осындай жүйелік белгілерге сәйкес. Топтаманың алдыңғы жағындағы антецедент А үшін, ал топтаманың артқы жағындағы анафора R үшін. 1-кестеде көрсетілгендей мүмкіндік векторын ұсынамыз.

Кесте 1-анафоралық қатынасты шешудегі мүмкіндік векторлары

№	мүмкіндік векторы	Үйлесімділік функциясы
1	Синтаксистік құрлым	А және R Синтаксистік құрылымы бірдей немесе жоқ, Y иә, N жоқ
2	Өзгертілетін шектеу түрі	Өзгертілген шекті түрі А және R бірдей немесе жоқ, Y иә, N жоқ
3	Қашықтық	А және R бір сөйлемде, 1 иә, 2 интервал, 3 жоқ
4	Дара немесе көпше	А және R бірдей жекеше немесе көпше, Y иә, N әртүрлі, U жоқ
5	Гендерлік категория	А және R гендерлік категориясы бірдей Y бірдей, N әртүрлі ретінде, U жоқ
6	Түр	А және R бірдей нысан түрі .Y бірдей, N әртүрлі ретінде, U жоқ
7	Сәйкестік туралы ақпарат	А және R толық сәйкестігі. Y иә, N әртүрлі
8	Сөздердің ұқсастығы туралы ақпарат	А және R нобайы сәйкес .Y иә, N әртүрлі
9	Қысқартылған сөздер	А және R қысқартылған сөздер.Y иә, N әртүрлі

Тірек векторлары машиналардың моделіне негізделген іске асыру

1) Оқыту деректерін масштабтау. LibSVM - бұл векторлық машиналарды қолдауға арналған кітапхана, әрбір деректер данасы нақты сандар векторы ретінде ұсынылуын талап етеді. Демек, категориялық атрибуттар болса, біз оларды біріншісінде сандық деректерге түрлендіруіміз керек. m -санатының төл сипатын көрсету үшін m санын пайдалануды ұсынамыз. m санының біреуін бір, ал басқаларын нөл деп есептейміз. Мысалы, қызыл, жасыл, көк сияқты үш санатты атрибутты $(0,0,1)$, $(0,1,0)$ және $(1,0,0)$ ретінде көрсетуге болады. SVM қолданбас бұрын оларды масштабтау өте маңызды. Негізгі артықшылығы - үлкен сандық диапазондағы атрибуттардың үстемділігін болдырмау. Тағы бір артықшылығы - есептеу кезінде сандық қиындықтарды болдырмау. Себебі, ядро мәндері әдетте сызықтық ядро және көпмүшелік ядро сияқты сипаттамалық векторлардың ішкі көбейтіндісіне тәуелді болғандықтан, атрибуттардың үлкен мәндері сандық мәселелерді тудыруы мүмкін. Әр атрибутты $[-1; +1]$ немесе $[0; 1]$ ауқымына сызықтық түрде масштабтауды ұсынамыз.

2) оқыту параметрін оңтайландыру. Іздеу параметрлерінде біз Libsvm, Python, Gnuplot үш құралын қолдандық. Python gnuplot бағдарламасын орнатқаннан және орнатқаннан кейін сіз параметрлерді автоматты түрде іздеуді жүзеге асыра аласыз. 2-суретте оңтайлы параметрлерді табу үшін Python Пайдалану процесі көрсетілген.

Кесте 2- Эксперименттік нәтижелерін салыстыру

Эксперимент	Жіктеу даналары	Дәлдік нәтижелері
максималды энтропия әдісі	Оң даналары	74%
	Теріс даналары	79%
SVM	Оң даналары	76%
	Теріс даналары	82%
Оқу деректерін масштабты өңдеу және оқыту параметрлерін оңтайландыру эксперименті	Оң даналары	80%
	Теріс даналары	83%

Кестеден эксперименттік деректерді салыстыра отырып, жіктеу үшін максималды энтропиямен жіктеу алгоритмі және тірек векторларының машиналары қолданылғанын көреміз, екі әдістің нәтижесінде айтарлықтай айырмашылықтар бар, бірақ үшінші экспериментте деректерді масштабты өңдеуді және оқыту параметрлерін оңтайландыру нәтижелерінде айқын жақсартуларға ие. бұл сонымен қатар деректерді өңдеудің оң нәтиже беретінін және өте тиімді екенін түсіндіреді. Жоғарыда келтірілген эксперименттік деректерден оң даналардың нәтижелері теріс даналардың нәтижелерінен жақсырақ екенін анық көруге болады, сонымен қатар біз статистикалық талдау арқылы теріс даналардың саны оң даналардың санынан көп екенін оңай табамыз.

Талқылау. Зерттеу жұмысы көрсеткендей, қазақ тіліндегі анафоралық қатынасты шешуде тірек векторлық машина арқылы шешу тиімділігі жоғары екендігін, машиналық оқытуда тірек векторлық машиналарды(SVM) және Libsvm кітапханасын қолданылуы табиғи тілді өңдеуде нәтижесін жақсартудағы тиімділігін көрсетті. Бұл зерттеу жұмысымызда ең басты артықшылықтар тірек векторлық машинаны(SVM) қолданылған кезінде сынақ-оқу деректерін масштабтау, оқыту параметрін оңтайландыру, сөздік белгілер, сөздер арасындағы қашықтық, жүйелілік белгілер және т.б. пайдаланылды. Оқу деректерін масштабты өңдеу және оқыту параметрлерін оңтайландыру эксперименті жасалынып максималды энтропия әдісі мен SVM әдісімен салыстырғанда әлде қайда жақсы нәтиже алынды(2-кестеде көрсетілген).

Қазақ тіліндегі есімдік анафорасын шешу бойынша жасалынған зерттеу жұмыстармен (Қалман т.б., 2022:125) салыстырғанда әлде қайда жақсы нәтижелер алынды.

Қортынды. Бұл мақалада тірек векторлық машиналар моделіне негізделген анафораны шешу модельі ұсынылды. модельді құру және эксперименттер жүргізу үшін сөздік белгілер, сөздер арасындағы қашықтықты, дәйектілікті және басқа сипаттамаларын жан-жақты қарастырады.

Модель жұмысын тестілеу мақсатында деректерді оқу жүргізілді. Сынақ деректері жинағы ретінде Tengrinews.kz- тен жаңалықтар топтамасы және F. Мұстафин әңгімелерінен үзінділерді қолдандық. оқыту параметрін оңтайландыруда Libsvm, Python, Gnuplot үш құралын қолдандық және модельді

жүзеге асыру кезінде support vector machines toolkit құралдар жиынтығы қолданылды.

Анафоралық қатынастың шешілу дәлдігін бағалау мақсатында эксперименттер жасалды және нәтижесін талдау мақсатында кесте түрінде алынды (кесте 2). Таңдалған белгілер екілік жіктеу құрылымына шешуші әсер ететіндіктен, біз зерттеу барысында екілік жіктеуге байланысты көптеген мәселелердің бар екенін ескердік.

Эксперименттер нәтижесінде анафоралық қатынасқа деректерді өңдеу және масштабтау тиімді болғанын және тірек векторлық машиналар арқылы анафоралық қатынасты шешу жақсы нәтиже бергенін көреміз.

Information about the authors:

Kalman Gulzhamal – graduate student at the Department of Information Systems, L.N. Gumilyov Eurasian National University, E-mail: guljamal14@gmail.com, DOI:<https://orcid.org/0000-0001-8863-9447>;

Sambetbayeva Madina Aralbayevna – Ph.D., Associate Professor, Nur-Sultan, Kazakhstan, Department of Information Systems, Eurasian National University named after L.N. Gumilyov. Senior Researcher, Institute of Information and Computing Technologies, Almaty, Kazakhstan, E-mail: madina_jgtu@mail.ru, <http://orcid.org/0000-0001-9358-1614>;

Aktaeva Dilara Aidosovna – Master of technical sciences, Department of Information Systems, Abay Myrzakhmetov Kokshetau University, Kokshetau, Kazakhstan, E-mail: dilaraabilmazinova@gmail.com, <https://orcid.org/0000-0002-7307-2481>;

Ilyubayev Adelzhan – Master of technical sciences, Kokshetau, Kazakhstan department of information systems and computer science, A. Myrzakhmetov Kokshetau university. Tel. +7776104747. E-mail: adik_better@mail.ru , DOI: <https://orcid.org/0000-0001-6333-0718>.

ӘДЕБИЕТТЕР:

Абрамов В.Е., Абрамова Н.Н., Некрасова Е.В., Росс Г.Н. (2011) Статистический анализ связности текстов по общественно-политической тематике // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: тр. 13 Всерос. науч. конф. Воронеж, с 127–133. (орыс тілінде).

Қалман Г., Самбетбаева М.А., Жұмабай Е.С. (2022) Қазақ тіліндегі есімдік анафорасын шешу алгоритімі, Инновациялық Еуразия университетінің Хабаршысы. № 2 ISSN 2709-3077, 126 б. (қазақ тілінде).

Толпегин П.В. (2006) Новые методы и алгоритмы автоматического разрешения референции местоимений третьего лица русскоязычных текстов. Ком-Книга, – 88 с. (орыс тілінде).

Толпегин П.В., Ветров Д.П., Кропотов Д.А. (2006) Алгоритм автоматизированного разрешения анафоры местоимений третьего лица на основе методов машинного обучения. Международной конференции «Диалог 2006» – Компьютерная лингвистика и интеллектуальные технологии – Бекасово, с 504–507. (орыс тілінде).

Chang C.C., Lin C.J. (2014). LIBSVM-ALibraryforSupport Vector Machines, availableat [Электронныйресурс]. – Режимдоступа: www.csie.ntu.edu.tw/~cjlin/libsvm.

Elango P. (2006) Coreference Resolution: A Survey. Technical Report. UW-Madison. (ағылшын тілінде).

Jun Lang, Bing Qin, Ting Liu, Sheng Li (2007). Intra-document Coreference Resolution: The state of the art. *Journal of Chinese Language and Computing*, pp. 227-253. (ағылшын тілінде).

Kibrik A.A. (1996), Anaphora in Russian narrative discourse: A cognitive calculative account in B, Fox (ed.) *Studies in anaphora*, Amsterdam, pp. 255–304. (ағылшын тілінде).

Kibrik A.A., Dobrov G.B., Khudyakova M.V., Loukachevitch N.V., Pechenyj A. (2013), A corpus-based study of referential choice: Multiplicity of factors and machine learning techniques, *Text processing and cognitive technologies. Cognitive modeling in linguistics: Proceedings of the 13th International Conference, Corfu*, pp. 118–126. (ағылшын тілінде).

Mitkov R. (1999) Anaphora resolution: the state of the art. *Citeseer*. (ағылшын тілінде).

Mitkov R. (2003) Anaphora resolution. *The Oxford handbook of computational linguistics*, ch.14, N.Y.: Oxford university. (ағылшын тілінде).

Prokofyev R., Tonon A., Luggen M., Vouilloz L., Difallah D.E., Cudr'e-Mauroux P. (2015) SANAPHOR: Ontology-Based Coreference Resolution. *14th International Semantic Web Conference, part I, LNCS, vol. 9366*, pp. 458-473. (ағылшын тілінде).

Qiu Rong sheng, Dong Yunjie. (2003) Research developments of decomposition algorithm for SVM QP problem[J]. *Pattern Recognition and Artificial Intelligence*. vol.16, pp 63-69. (ағылшын тілінде).

Rong-En Fan, Pai-Hsuen Chen, Chih-Jen Lin. (2005) Working Set Selection Using Second Order Information for Training Support Vector Machines[J]. *Journal of Machine Learning Research* pp 889-1918. (ағылшын тілінде).

Zhou, H., Li, Y., Huang, D., Zhang, Y., Wu, C., Yang, Y. (2011). *Combining syntactic. Weka 3: Data Mining Software in Java*, available at University of Waikato. Retrieved from: www.cs.waikato.ac.nz/ml/weka/ 20. (ағылшын тілінде).

REFERENCES:

Abramova N.N., Abramov V.E., Nekrasova E.V., Ross G.N. (2011), Statistic analysis of social and political texts coherence [Statisticheskij analiz svjaznosti tekstov po obshchestvenno-politicheskoy tematike], *Proceedings of the 13th All-Russian Scientific Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections” [Trudy 13j Vserossijskoj nauchnoj konferentsii “Elektronnyye biblioteki: perspektivnye metody i tehnologii, elektronnyye kollekcii”]*, Voronezh, pp. 127–133. [in Russian].

G. Kalman, M.A. Sambetbayeva, Y.S. Zhumabay (2022) Algorithm for solving the anaphora of a pronoun in the Kazakh language, *bulletin of the innovative university of Eurasia*, № 2 ISSN 2709-3077, pp 126 (in Kazakh).

Tolpegin P.V. (2006), The new methods and algorithms of automated third person pronominal reference resolution of Russian texts, [Novye metody i algoritmy avtomaticheskogo razresheniya referentsii mestoimenij tret'ego litsa russkojazychnyh tekstov], *Komkniga*, p-88. [in Russian].

Tolpegin P.V., Vetrov D.P., Kropotov D.A. (2006), Automated third person anaphora resolution algorithm based on machine learning methods [Algoritm avtomatizirovannogo razresheniya anafory mestoimenij tret'ego litsa na osnove metodov mashinnogo obuchenija], *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2006”*, Bekasovo, pp. 504–507. [in Russian].

Elango P. (2006) Coreference Resolution: A Survey. Technical Report. UW-Madison. (in Eng.).

Jun Lang, Bing Qin, Ting Liu, Sheng Li. Intra-document Coreference Resolution: The state of the art [J]. *Journal of Chinese Language and Computing*, 2007, 17(4):227-253. (in Eng.).

Kibrik A.A. (1996), Anaphora in Russian narrative discourse: A cognitive calculative account in B, Fox (ed.) *Studies in anaphora*, Amsterdam, pp. 255–304. (in Eng.).

Kibrik A.A., Dobrov G.B., Khudyakova M.V., Loukachevitch N.V., Pechenyj A. (2013), A corpus-based study of referential choice: Multiplicity of factors and machine learning techniques, Text processing and cognitive technologies. Cognitive modeling in linguistics: Proceedings of the 13th International Conference, Corfu, pp. 118–126. (in Eng.).

Mitkov R. (2003) Anaphora resolution. The Oxford handbook of computational linguistics, ch.14, N.Y.: Oxford university. (in Eng.).

Prokofyev R., Tonon A., Luggen M., Vouilloz L., Difallah D.E., Cudr'e-Mauroux P.(2015) SANAPHOR: Ontology-Based Coreference Resolution.14th International Semantic Web Conference, part I, LNCS, vol. 9366, pp. 458-473. (in Eng.).

Qiu Rong sheng, Dong Yunjie. Research developments of decomposition algorithm for SVM QP problem[J]. Pattern Recognition and Artificial Intelligence.vol.16, March.2003.63-69. (in Eng.).

Rong-En Fan, Pai-Hsuen Chen, Chih-Jen Lin. Working Set Selection Using Second Order Information for Training Support Vector Machines[J]. Journal of Machine Learning Research 6(2005)1889- 1918. (in Eng.)

Zhou, H., Li, Y., Huang, D., Zhang, Y., Wu, C., Yang, Y. (2011). Combining syntactic. Weka 3: Data Mining Software in Java, available at University of Waikato. Retrieved from: www.cs.waikato.ac.nz/ml/weka/ 20.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 68-80

<https://doi.org/10.32014/2022.2518-1726.157>

МРНТИ81.93.29

УДК 004.056.5

**С.Т. Мамбетов^{1*}, Е.Е. Бегимбаева^{1,2}, С.К. Джолдасбаев³, Б.О. Куламбаев³,
Г.Н. Казбекова⁴**

¹Казахский национальный университет им. аль-Фараби, Алматы, Казахстан;

²Казахский национальный исследовательский технический университет
имени К.И. Сатпаева, Алматы, Казахстан;

³Международный университет информационных технологий,
Алматы, Казахстан;

⁴Международный казахско-турецкий университет имени Ходжи Ахмеда
Ясави, Туркестан, Казахстан.

E-mail: mambetov.saken@gmail.com

О МОНИТОРИНГЕ УГРОЗ И УЯЗВИМОСТЕЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Аннотация. Стремительная информатизация в разных сферах деятельности общества имеет радикальное влияние на доступ и обработку информации. В буквальном смысле за последние два десятилетия появились новые формы транслирования и передачи данных и совершенно новые виды деятельности внутри мировой Интернет-паутины, связанные с последними новшествами. Широко распространилось понятие «социальная сеть», в котором люди без всяких страхов и не имея представления о риске афишируют свои данные на публику. Сейчас крайне редко можно встретить человека, который не пользуется ни одним приложением социальной сети, и чаще всего верно утверждение, что один человек пользуется несколькими разными приложениями. Кроме того, в ходе работы используются Интернет-сервисы, поисковые системы, электронная почта. Соответственно, возрастают и появляются новые, ранее неизвестные виды разного рода угроз и уязвимости информационных систем. С развитием Интернета атаки хакеров также были сосредоточены на тех же уязвимостях и новых типах угроз. Система мониторинга помогает выявлять эти уязвимости и угрозы. В свою очередь, мониторинг ситуации с

информационной безопасностью создает возможность контролировать работу ИТ-ресурсов, сетевых программ, устройств и веб-сервисов любой компании в автоматическом режиме. Благодаря постоянному мониторингу (контролю) можно вовремя выявить уязвимости и угрозы информационной системы и остановить их работу. Своевременный мониторинг снижает размер ущерба, причиняемого информационной системе. В данной статье мы приводим анализ по направлению исследования мониторинга угроз и уязвимостей информационной системы, методов защиты от угроз с приоритетом, основанных на анализе данных тематических Интернет-ресурсов. Кроме того, будет сделан обзор предыдущих статей авторов.

Ключевые слова: мониторинг; Интернет-угрозы; уязвимости; кибербезопасность.

**С.Т. Мамбетов^{1*}, Е.Е. Бегимбаева^{1,2}, С.К. Джолдасбаев³, Б.О. Куламбаев³,
Г.Н. Казбекова⁴**

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;

²Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті,
Алматы, Қазақстан;

³Халықаралық Ақпараттық технологиялар университеті, Алматы, Қазақстан;

⁴Қожа Ахмет Яссауи атындағы Халықаралық Қазақ-Түрік университеті,
Түркістан, Қазақстан.

E-mail: *mambetov.saken@gmail.com*

АҚПАРАТТЫҚ ЖҮЙЕНІҢ ҚАУШТЕРІ МЕН ОСАЛ ТҰСТАРЫНЫҢ МОНИТОРИНГІ ТУРАЛЫ

Аннотация. Қоғамның әртүрлі салаларындағы жедел ақпараттандыру ақпаратқа қол жеткізу мен өндеуге түбегейлі әсер етеді. Соңғы екі онжылдыққа инновациялармен байланысты ғаламдық интернет желісінде хабар тарату мен деректерді берудің жаңа нысандары және жаңа қызмет түрлері пайда болды. Әлеуметтік желі ұғымы кең тарап, онда адамдар қорқынышсыз, қауіп туралы ешқандай түсініксіз өз деректерін көпшілікке жариялап отырады. Қазір ешқандай әлеуметтік желі қосымшасын пайдаланбайтын адам өте сирек кездеседі және көбінесе бір адам бірнеше түрлі қосымшаларды пайдаланады деген мәлімдеме шындыққа сәйкес келеді. Оған қоса жұмыс барысында интернет сервистерді, іздеу жүйелерін, электронды пошталарды қолданады. Сәйкесінше, ақпараттық жүйелердің жаңа, бұрын белгісіз, әртүрлі қауіптер мен осал түрлері көбейіп, пайда болуда. Ғаламтор желісінің дамуымен хакерлердің де шабуылдары сол осалдықтар мен жаңа қауіп түрлеріне бағытталды. Осы осалдықтар мен төнетін қауіптерді анықтауға мониторинг жүйесі көмектеседі. Өз кезегінде ақпараттық қауіпсіздік жағдайының мониторингі кез-келген

компанияның IT-ресурстарының, желілік бағдарламаларының, құрылғылар мен веб-сервисінің автоматты режимде жұмысын бақылауға мүмкіндік туғызады. Сондай тұрақты мониторинг (бақылау) арқасында ақпараттық жүйенің осалдықтары мен қауіптерін дер кезінде анықтап, олардың жұмысын тежеуге мүмкіндік береді. Уақытылы бақылау жүргізілген кезде ақпараттық жүйеге келтірілетін шығын көлемін азайтады. Бұл мақалада біз тақырыптық интернет-ресурстар деректерін талдау негізінде ақпараттық жүйенің қауіп-қатерлері мен осал тұстарының мониторингін, қауіптерден басымдықпен қорғау әдістерін зерттеу бағыты бойынша талдауды ұсынамыз. Соған қоса алдыңғы авторлардың мақалаларына шолу жасалады.

Түйін сөздер: мониторинг; интернет қауіптер; осалдықтар; киберқауіпсіздік.

S.T. Mambetov^{1*}, Ye. Ye. Begimbayeva^{1,2}, S. Joldasbayev³, B.O. Kulambayev³, G.N. Kazbekova⁴

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan;

²Satbayev University, Almaty, Kazakhstan;

³International University of Information Technologies, Almaty, Kazakhstan;

⁴Akhmet Yassawi International Kazakh-Turkish University,
Turkestan, Kazakhstan.

E-mail: *mambetov.saken@gmail.com*

ABOUT MONITORING THREATS AND VULNERABILITIES OF THE INFORMATION SYSTEM

Abstract. The rapid informatization in various areas of society has a radical impact on the access and processing of information. In the literal sense, over the past two decades, new forms of broadcasting and data transmission have appeared and completely new types of activities within the global Internet network associated with the latest innovations. The concept of the Social Network has spread widely, in which people, without any fears and without any idea of the risk, advertise their data to the public. Now it is extremely rare to find a person who does not use any social network application, and most often the statement is true that one person uses several different applications. In addition, Internet services, search engines, and e-mail are used in the course of work. Accordingly, new, previously unknown, types of various threats and vulnerabilities of information systems are increasing and appearing. With the development of the Internet, hacker attacks have also focused on the same vulnerabilities and new types of threats. The monitoring system helps to identify these vulnerabilities and threats. In turn, monitoring the information security situation makes it possible to monitor the operation of IT resources, network programs, devices and web services of any company in automatic mode. Thanks to

constant monitoring (control), it is possible to identify vulnerabilities and threats of the information system in time and stop their work. Timely monitoring reduces the amount of damage caused to the information system. In this article, we provide an analysis in the direction of the study of monitoring threats and vulnerabilities of the information system, methods of protection against threats with priority, based on the analysis of data from thematic Internet resources. In addition, a review of the authors' previous articles will be made.

Key words: monitoring, Internet threats, vulnerabilities, cybersecurity.

Введение. Сегодня глобальное развитие проходит эру цифровизации, и если одни сообщества, регионы и даже целые страны уже полностью перешли на новый уровень, в котором большинство услуг производятся в цифровом пространстве, то некоторые группы только переходят к этому направлению. Никому не секрет, что цифровые технологии являются передовыми в эпоху, которую мы переживаем, и данное внедрение положительно сказывается в первую очередь на экономическом росте, обеспечивая высокую производительность, хранение, обработку и транспортировку цифровых ресурсов. В Республике Казахстан, с целью ускорения темпов развития экономики, улучшения качества жизни населения за счет цифровых технологий, также реализуется программа «Цифровой Казахстан» (Государственная программа «Цифровой Казахстан», 2018), где одним из важных пунктов указано обеспечение информационной безопасности в сфере ИТ. Еще одним катализатором послужил факт распространения COVID-19, когда практически более 80% (Gurova, 2020) офисных работников крупных корпораций перешли к дистанционной форме и после ослабления карантинных мер в целях безопасности долгое время оставались на удаленной работе. В любом случае после таких глобальных изменений мир уже не будет прежним.

Соответственно, при оптимизации процессов обработки различного рода данных необходимо обеспечить надежное хранение и оперативный доступ для участников информационного обмена, где существуют риски утечки, нарушения целостности и другие подобные уязвимости. Следует также брать во внимание, что с количеством роста пользователей также растет и количество уязвимости информационного пространства, так как это расширяет возможности злоумышленников, повышает вероятность непредвиденных случаев и рассеивает возможности их предотвращения устоявшимися методами. В данной работе проводится анализ типов угроз и уязвимостей, анализ методов и программных реализации предотвращения и борьбы с угрозами и уязвимостями информационной системы. При анализе информационной безопасности информационной системы организаций и предприятий проводится рассмотрение нескольких способов при работе пользователей общей системы с данными, таких как сбор и обработка данных из открытых источников, мониторинг распределения ресурсов и способы повышения безопасности.

Материалы и методы исследования. Обнаружение уязвимостей информационной системы дает преимущество для предотвращения разного рода действий злоумышленников. Чтобы бороться с проблемой, для начала нужно определить характер проблемы. Обычно программисты создают большое количество разного рода методы, алгоритмы, программы и даже целые системы защиты от взлома, тем не менее они происходят. Злоумышленник планирует атаку подобно шахматной игре, предполагая вопросы, которые цель атаки может задать, так что у него могут быть готовы подходящие ответы (Mitnick et al., 2002). Мы не стали описывать причины действий злоумышленников в данной работе, хотя, как известно, в большинстве случаев мотив раскрывает характер действий, тем не менее, мы попытались определить виды угроз, которые распространены в сети и те, которые приносят большие убытки.

По данным исследовательского центра Data Pro Research (США), в 62% причинами утечки или утраты информации служит человеческий фактор, среди которого 16% – повреждение софта, 16% – хищение персональной информации, 12% – подмена и фальсификация информации, 54% – финансовые махинации разного рода: переводы или кража с электронного счета, оплата услуг за чужой счет и т.д. Также прогнозируется, что если количество киберугроз будет расти и дальше, то объем расходов на решения в области кибербезопасности к 2022 году достигнет 133,7 миллиардов долларов США (Data Pro Research, 2022).

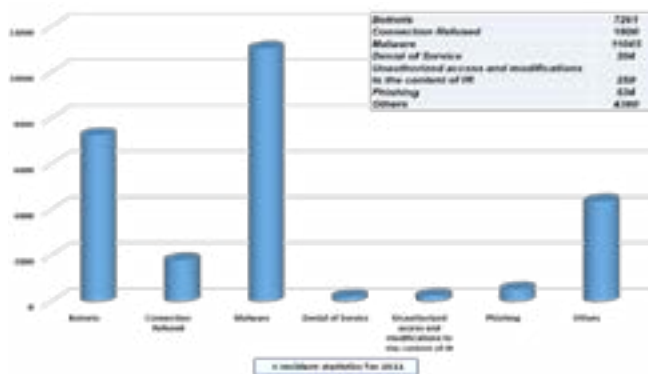


Рисунок 1. Статистика инцидентов киберугроз на территории РК за 2021 год

Казахстан 2017 году в рейтинге 194 стран по уровню кибербезопасности Global Cybersecurity Index был на 82 месте, а в 2021 году занял 31 место из 182 возможных (Global Cybersecurity Index, 2021), что показывает рост интереса к кибербезопасности в стране и актуальность исследований в этой области. Потому что информационная безопасность является задачей, которая требует регулярного процесса мониторинга уязвимостей и не имеет конечного решения еще со времен появления первых компьютерных сетей. На рисунке-1 приведена статистика, взятая из сайта о кибератаках, совершенных в Казахстане за 2021 год.

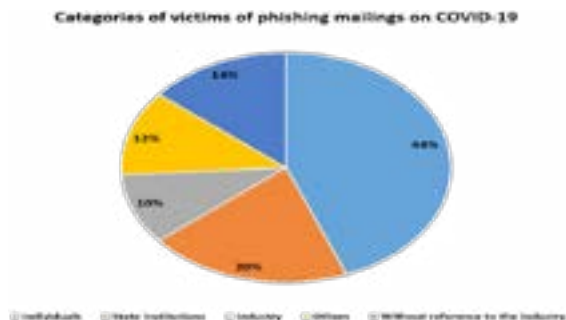


Рисунок 2. Категории жертв фишинговых рассылок на тему COVID-19

Следует взять за внимание и тот факт, что мошенничество не всегда осуществляется в физическом мире, существуют разновидности онлайн-форм, где приманки включают в себя привлекательную рекламу, которая ведет на вредоносные сайты или побуждает пользователей загрузить приложение, зараженное вредоносным ПО. Так, на примере в 2020 году злоумышленники подхватили тему всеобщего беспокойства по поводу COVID-19 и стали использовать ее для фишинговых писем (рисунок 2).

В (Миков, 2014) анализируются возможности методов и средств оценки рисков информационной безопасности применительно к различным этапам процесса оценки и построена схема процесса в виде вложенных алгоритмов с указанием взаимосвязей между всеми этапами, также рассмотрены наиболее эффективные способы реализации этих этапов. Автор предлагает при оценке рисков информационной безопасности использовать не один универсальный метод, а комбинировать методы и средства на разных этапах процесса: DFD или IDEF0 – на этапе анализа потоков данных в информационной системе, экспертный опрос с методами Дельфи, конкордации или симплекс-методом – на этапе оценки факторов риска, методы искусственного интеллекта, среди них гибридные модели – на этапе оценки риска, соответствующие математические методы – на этапе экономической оценки защиты информации и минимизация, принятие, передача или уклонение от риска – на этапе реализации управления рисками.

В (Киселева и др., 2017) определено понятие информационного риска, рассмотрены методы для оценки уровня текущего состояния информационной безопасности предприятия, для получения точных удовлетворительных результатов оценки предложено использовать комплексный подход к оценке рисков на основе существующих методик. Также в работе (Аникин и др., 2015) рассмотрены основные методы качественной и количественной оценки рисков информационной безопасности в корпоративных информационных сетях и поднята актуальность применения теории нечетких множеств при оценке рисков информационной безопасности. Значимым для нас считается, что здесь приведены методы, разработанные IT специалистами разных агентств разных стран.

В (Xiong et al., 2019) представлен обзор моделирования угроз на основе систематических запросов в четырех ведущих научных базах данных, и было оценено 176 статей, разделенных три отдельных кластера: (1) работы, вносящие вклад в моделирование угроз, (2) статьи, использующие существующий подход к моделированию угроз, и (3) вводные статьи, представляющие работу, связанную с процессом моделирования угроз. Авторы верно подметили, что большая часть работы по моделированию угроз по-прежнему выполняется вручную, и существует ограниченная уверенность в их валидации.

В (Russo et al., 2019) была разработана CYber Risk Vulnerability Management (CYRVM) – для упрощения и улучшения автоматизации и непрерывности оценки кибербезопасности. Основными нововведениями CYRVM являются сочетание в единой и простой в использовании программной веб-платформе онлайн-инструмента оценки уязвимостей в рамках структуры анализа рисков в соответствии с NIST 800-30 Risk Management руководящие принципы и интеграция прогнозных решений, способных предложить пользователю рейтинг и классификацию риска.

В (Chhajed et al., 2014) приводится обзор некоторых алгоритмов и методов обнаружения межсайтового скриптинга (XSS), а также sniffинга (sniffing) контента, когда в ход вступает изменение данных в содержимом текстовых, pdf-файлов и изображений.

В работе (Ajmal et al., 2021) хорошо описано положение, когда злоумышленники стремятся скомпрометировать организации и их данные с помощью передовых скрытых методов, используя законные инструменты, а также предлагается метод по определению системе незнакомых угроз, на новом подходе к моделированию имитации противника (отображение каждой соответствующей фазы) внутри подхода поиска угроз с минимальными ресурсами. Предлагаемый подход может быть использован для разработки атакующей среды, ориентированной на безопасность, в которой организации могут выявлять расширенные механизмы атак и проверять их способность обнаруживать атаки. Еще одна работа, на которую следует обратить внимание (Bakić et al., 2021), где был оценен прогресс в области кибербезопасности за последние 10 лет, спровоцированная в свою очередь в результате атаки Stuxnet в 2010 году. Авторы предлагают реактивные и упреждающие меры для снижения вездесущих рисков кибербезопасности.

В (Laković et al., 2021) рассматриваются вопросы кибербезопасности: существующие модели кибербезопасности и их совершенствования, предлагается методология семантического расширения и улучшения моделей кибербезопасности. В работе (Velicković et al., 2021) рассматривается проблема инфодемии – переизбыток информации о проблеме, обычно ложной и непроверенной, вызванная объявлением пандемии COVID-19. Был сформирован список веб-сервисов, которые предоставляют достоверные данные о пандемии из соответствующих источников и как таковые могут использоваться в борьбе с COVID-19 инфодемией, описаны данные COVID-19 службы

Freemium Web API с набором методов GET. На основании полученных результатов был предложен способ использования данного веб-сервиса в борьбе с инфодемией COVID-19.

В (Болатбек, т.б., 2022) этой работе поднимается проблема кибербезопасности, которая в настоящее время очень важна из-за некоторых угроз безопасности и кибератак. Дается определение кибербезопасности и приводятся наиболее распространенные виды атак на сегодняшний день. Кроме того, было показано, что некоторые проблемы кибербезопасности можно решить с помощью методов NLP обработки естественного языка, и был проведен систематический литературный обзор текущих работ.

Также следует дать описание экспертным системам, которые предназначены для решения классификационных задач в определенной области на основе базовых знаний, полученных путем опроса квалифицированных и представленных правилами системы классификации. В системах безопасности информационных технологий экспертные системы используются в интеллектуальных системах защиты информации на основе стенографической модели и содержат, как правило, неявные операторы преобразования в командный код каждого приложения отдельно, где имеются конкретные файловые системы и прикладные программное обеспечения отдельно. Возможность брать за основу опыт экспертов и информационную безопасность в виде правил являются преимуществами таких систем, и процесс аналогичен работе человеческой логики. Процесс описания последовательности правил определения реализовывается путем прямого и обратных операций (рисунок 4).



Рисунок 3. Архитектура простой экспертной системы мониторинга угроз и уязвимостей

Говоря об экспертной системе, стоит упомянуть, что для больших объемов данных и при частом обновлении контента обычно предлагается парсинговый метод – метод синтаксического анализа. Алгоритм действий такой программы приведен на рисунке-4.

Недостатком систем является требование больших вычислительных мощностей готовых правил без обоснования необходимости для конкретной цели. Учитывая необходимость обеспечения скрытой и своевременной передачи данных в современных экспертных системах, можно предположить,

что использование технологии самомодификации файлов, а именно программы скрытой перекодировки, важно для разработки технологий в области искусственного интеллекта (Shterenberg et al., 2016).



Рисунок 4. Алгоритм действий парсинговой программы

Использование машинного обучения в качестве решения также пользуется большим успехом. Они подразумеваются в двух видах в зависимости от управления: контролируемые, неконтролируемые.

Результаты исследования. Таким образом, был выявлен ряд решений отслеживания угроз, которые применимы в информационных системах. Как показывает практика, системы мониторинга угроз условно можно разделить на категории:

1. Системы анализа в реальном времени;
2. Системы сбора данных действий пользователей;
3. Системы детекции аномалии действий пользователей;
4. Системы контроля внутренних операции;
5. Системы поиска и выявления угроз, ориентированные на улучшение системы безопасности.

Системы анализа в реальном времени позволяют выполнять обработку поступающих через канал данных и анализировать каждое сообщение по мере его получения. Также анализ в реальном времени преобразует определенные данные, геозон и обнаружения произошедших инцидентов и по завершению создаются набор выходных данных, которые будут храниться в векторном слое и также отправляться в виде оповещений.

В ходе исследования были рассмотрены некоторые системы анализа в реальном времени. Одним из таких систем является информативно-аналитический Интернет-сервис Mindscan, который имеет графический интерфейс в качестве Web-браузера, развертывается в облаке SaaS, и в основном занимается мониторингом и анализом социального Медиа.

Данный программный сервис ориентирован на крупные и средние организации и доступна не каждому пользователю, так как имеет плату за оказание услуг. Еще одним широко распространенным продуктом является система онлайн-мониторинга *Katyusha*. Данная система подходит как для среднего персонального компьютера, так и для смартфонов *iOS* и *Android*, с возможностями показа новостей в режиме реального времени, информативной аналитики и систематического выявления информативных атак и оценки эффективности. Также данная система выполняет поиск необходимой информации. Облачная система позволяет оценивать степень распространения той или иной определенной информации и перепубликаций одного источника. Система комплексного управления с анализом новостей для СМИ и социальных сетях в режиме реального времени *SCAN*, которая является профессиональным инструментом специалистов связи с общественностью. Данная система имеет возможность сбора данных более чем из 20 000 источников информации, в числе которых *Facebook*, *VK*, *Twitter* и многие другие более распространенные социальные сети. Существуют также большинство других не менее распространенных систем мониторинга и анализа информативных сред, таких как *SemanticForce*, *Cision*, *Awario*, *PressIndex*, *YouScan*, *Seldon.News*, *Mediascope*, *Mention* и множество других (Базенков и др., 2013).

Что касается сбора данных действий пользователей браузер пользователя передает в систему аналитики информацию о действиях, выполняемых им на сайте. Сайт в свою очередь, также может передавать данные как напрямую в систему аналитики, так и в интегрированную CRM-систему. CRM, используя полученные от сайта данные, может передавать их сразу в систему аналитики или же после их предварительной обработки. Несмотря на множество источников, из которых данные могут поступать в систему аналитики, наиболее исчерпывающими являются те данные, которые передаются в систему непосредственно браузером, поскольку не были подвержены модификации. Для сбора пользовательских данных при открытии сайта на страницу встраивается код, который непосредственно выполняет сбор данных. С помощью этого можно отслеживать передвижения клиента по вашему сайту и строить граф переходов, определять наиболее частые действия того или иного пользователя (Gubanov, 2020).

Системы детекции аномалии действий пользователей было определено в списке как отдельный пункт, так как бывают ситуации, когда эффект действия будет распознан позже, чем само действие. К примеру, была произведена определенная операция пользователем, а обработка и ответ из сервера проходит через большое количество времени. Таким образом, при возобновлении операции браузер клиента отключен и возможно даже сайт заблокирован и нет возможности определить информацию об успешном завершении действий. Для данных о трафике веб-сайта эти риски не редкость, но они могут нанести полноценный ущерб. Потеря некоторой полезной информации

может привести к ложным обвинениям и, следовательно, к неэффективным неденежным решениям.

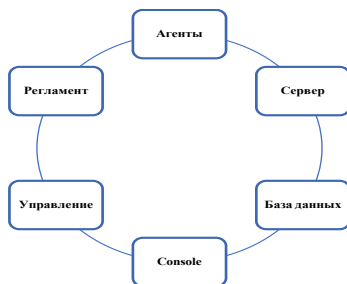


Рисунок 5. Основные компоненты для систем безопасности

Также были определены основные компоненты для систем безопасности (рисунок 5), где каждому компоненту определена своя роль: программные агенты собирают данные из различных источников, выше мы уже указывали программные продукты поддерживающие данный компонент; сервер анализирует поступающую информацию с заданными политикой и правилами; база данных консолидируют все события, сохраняет и производит быстрое развертывания необходимых данных в нужный момент; консоль управляет параметрами обработки и производит поправки в системе; управление производится системным администратором (или несколькими администраторами), которые работают с системой; и регламенты работы по мониторингу, выставляемые заказчиками.

Дискуссия. Поскольку 21 век – век информации, компьютерная индустрия получила широкое развитие. В частности, увеличилось количество пользователей Интернета. Однако увеличилось количество источников угроз из той же сети, а также увеличилось уязвимости информационных систем. Мы убеждены в необходимости мониторинга в целях предотвращения уязвимостей информационной системы и угроз системе. Система мониторинга позволит выявить и оценить слабые места информационной системы. Оценивание уязвимости позволяет нормально, без угроз работать системе. Жизненный цикл системы мониторинга состоит из 4 этапов.

- Выявление и профилактика активных поломок;
- Выявление и устранение потенциальных уязвимостей;
- Оценка риска угроз и уязвимостей информационной системы;
- Исправление уязвимостей и угроз информационных систем.

Закключение. Таким образом, было определены основные понятия систем анализа и мониторинга информативных интернет-ресурсов, их полезные и уязвимые качества. В работе были описаны эти системы мониторинга и безопасности, а также регламент, поддерживаемый на территории Республики Казахстан. Также был сделан вывод касательно того, что в большинстве случаев атаки злоумышленников на простых потребителей

приводятся в действие при помощи социальной инженерии. Следовательно, с появлением нового вида такой деятельности необходимо постоянно развивать систему мониторинга и применять новые методы защиты от подобных случаев, которые позволят улучшать защищенность информационной системы. Таким образом, в качестве метода защиты будет применен анализ данных по содержанию во избежание ложных данных и предотвращения, а также разработана информационная система с использованием методов искусственного интеллекта, которая повысит уровень безопасности данных, в первую очередь от уязвимостей, непреднамеренно допускаемых простыми пользователями (сотрудниками предприятия).

Благодарность. Авторы выражают благодарность научному руководителю проекта ГФ АР09259208 «Создание масштабируемой отказоустойчивой информационной системы цифровизации предприятия с использованием технологий Big Data» профессору ф.-м. н. Г.Т. Балакаевой.

Information about the authors:

Mambetov Saken Tolegenuly – Doctoral student of the Department of Information Systems of the Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: mambetov.saken@gmail.com; <https://orcid.org/0000-0002-7249-5378>;

Begimbayeva Yenlik Ye. – PhD, Associate Professor of the Department of Information Systems, Al-Farabi Kazakh National University, Almaty, Kazakhstan E-mail: enlik_89@mail.ru, <https://orcid.org/0000-0002-4907-3345>;

Joldasbayev Serik K. – Senior Lecturer of the Department of Computer Engineering, International IT University, Almaty, Kazakhstan, E-mail: serykjoldasbaev@mail.ru, <https://orcid.org/0000-0002-8689-1822>;

Kulambayev B.O. – Candidate of Technical Sciences, Associate Professor of the Department of Computer Engineering, International IT University, Almaty, Kazakhstan, E-mail: b.kulambayev@iitu.edu.kz, <https://orcid.org/0000-0001-8387-3736>;

Kazbekova G.N. – Candidate of Technical Sciences, Head of the Department of Computer Engineering, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan; E-mail: 3gulnur.kazbekova@ayu.edu.kz, <https://orcid.org/0000-0002-2756-7926>.

REFERENCES

Ajmal A.B., Shah M.A., Maple C., Asghar M.N., Islam S.U. (2021). Offensive Security: Towards Proactive Threat Hunting via Adversary Emulation. in IEEE Access, vol. 9, pp. 126023-126033, 2021, <https://doi.org/10.1109/ACCESS.2021.3104260> (in Eng.).

Anikin I.V., Emaletdinova L.Yu., Kirpichnikov A.P. (2015). Methods of information security risk assessment and management in corporate information networks. Bulletin of Kazan Technological University, 18 (6), 195-197. (in Russ.).

Bakić B., Milić M., Antović I., Savić D., Stojanović T. (2021). 10 years since Stuxnet: What have we learned from this mysterious computer software worm? 2021 25th International Conference on Information Technology, IT 2021, 9390103, <https://doi.org/10.1109/IT51528.2021.9390103> (in Eng.).

Bazenkov N.I., Gubanov D.A. (2013). Information systems for social networks analysis: a survey. Управление большими системами: сборник трудов, (41), 357-394. (in Russ.).

Bolatbek M.A., Bagitova K., & Mussiraliyeva Sh.Zh. (2022). A SYSTEMATIC REVIEW ON CYBERSECURITY ISSUES USING NATURAL LANGUAGE PROCESSING TECHNIQUES. Reports of NAS RK. Physico-Mathematical Series, (3), 52–70. <https://doi.org/10.32014/2022.2518-1726.139>.

Chhajed U., Kumar A. (2014). A Critical Review on Detecting Cross-Site Scripting Vulnerability. International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 4, April 2014 (in Eng.).

Data Pro Research (2021) [Electronic resource] (date of application 01/10/2022) URL: <https://dataprorsearch.com/research-cyber-security-2021.html> (in Eng.).

Global Cybersecurity Index 2020, Measuring commitment to cybersecurity (2021). [Electronic resource] (date of application 01/10/2022) URL: https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2021-PDF-E.pdf (in Eng.).

Gubanov I.A. (2020) An approach to simulation of user actions for data collection of analytics system. The electronic scientific journal "Young science of Siberia", no. 3(9). (in Russ.).

Gurova I.M. (2020) Remote Work as a Trend of Time: Results of Mass Testing. MIR (Modernization. Innovation. Research). 11(2):128-147. <https://doi.org/10.18184/2079-4665.2020.11.2.128-147> (in Russ.).

Kiseleva I.A., Iskajyan S.O. (2017). Information risks: methods of estimation and analysis. ITportal, (2 (14)), 10. (in Russ.).

Laković L., Ognjanović I., Šendelj R., Injac O. (2021). Semantically enhanced cyber security model for industry 4.0: Methodological framework. 2021 25th International Conference on Information Technology, IT 2021 this link is disabled, 2021, 9390120, <https://doi.org/10.1109/IT51528.2021.9390120> (in Eng.).

Mikov D.A. (2014). Analysis of methods and tools which are used in the various stages of information security risk assessment. Cybersecurity issues, (4 (7)), 49-54. (in Russ.).

Mitnick K.D., Simon W.L., Wozniak S. (2003). The Art of Deception: Controlling the Human Element of Security. John Wiley & Sons. (in Eng.).

On the approval of the State Program «Digital Kazakhstan» (2018) (date of application 01/05/2022) [Electronic resource]. URL: <https://adilet.zan.kz/rus/docs/P1700000827> (in Russ.).

Russo P., Caponi A., Leuti M., Bianchi G. (2019). A Web Platform for Integrated Vulnerability Assessment and Cyber Risk Management. Information, 10(7), 242. <https://doi.org/10.3390/info10070242> (in Eng.).

Shterenberg S.I., Kaflanov R.I., Druzhin A.S., Marchenko S.S. (2016) Method of use of self-modification files for secure communication in the expert system. H&ES Research. Vol. 8. No. 1. Pp. 71–75 (in Eng.).

Veličković Z.S., Veličković M.Z., Milivojević Z.N. (2021). Application of a Reliable Web API's in the Fight Against COVID-19 Infodemia. 2021 25th International Conference on Information Technology (IT). <https://doi.org/10.1109/IT51528.2021.9390128> (in Eng.).

Xiong W. & Lagerström R. (2019). Threat modeling – A systematic literature review. Computers & Security. <https://doi.org/10.1016/j.cose.2019.03.010> (in Eng.).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 81-92

<https://doi.org/10.32014/2022.2518-1726.158>

УДК 004.021

**У.Т. Махажанова^{1*}, Б. Тасуов², А.А. Муханова¹, А. Мухиядин¹,
Р.К. Жеткиншеков¹**

¹Л.Н. Гумилева атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

²М.Х. Дулати атындағы Тараз өңірлік университеті, Тараз, Қазақстан.

E-mail: makhazhan.ut@gmail.com

БҰЛДЫР ЖИЫНДАР ТЕОРИЯСЫ НЕГІЗІНДЕ БИЗНЕСТІҢ НЕСИЕ ҚАБІЛЕТІЛІГІН БАҒАЛАУ АЛГОРИТМІ

Аннотация. Шағын және орта бизнес көптеген елдердің экономикалық дамуында маңызды рөл атқарады. Нарықтық экономикалық жүйеде шағын кәсіпкерліктің қызмет етуі мен дамуының маңызды мәселелерінің бірі – қаржыландыру көздерін іздеу. Қазіргі шағын бизнес – бұл ашық және қарқынды дамып келе жатқан әлеуметтік-экономикалық және өндірістік-маркетингтік жүйе. Өндірістік-шаруашылық қызмет нәтижесінде алынған әртүрлі экономикалық, қаржылық, әлеуметтік және инвестициялық нәтижелерді кез келген бір көрсеткішті пайдалана отырып бағалау мүмкін емес. Көрсеткіштер жүйесін (кешенін) пайдалану қажет. Бұл белгілі көрсеткіштердің ешқайсысы шағын кәсіпорындардың шаруашылық қызметінің сан алуан түпкілікті экономикалық, қаржылық және әлеуметтік нәтижелерін көрсете алмайтындығына байланысты. Бұл мақалада шағын бизнесті несиелеудің ерекшеліктері қарастырылады. Шағын кәсіпкерлікті несиелеудің орындылығын анықтау үшін бағалау көрсеткіштерінің топтарының жиынтығы қарастырылды: салалық және аймақтық ерекшеліктер, шағын кәсіпкерлік қызметі және қаржылық-экономикалық көрсеткіштер. Бұлдыр жиындар теориясының математикалық аппаратын қолдану негізінде шағын және орта бизнестің несиелік қабілетін бағалау әдісі ұсынылған. Анық емес қорытынды әдісі интуиция мен тәжірибені модельге енгізуге мүмкіндік беретін бұлдыр ережелерді пайдаланады. Ережелер көрсеткіштер мен коэффициенттерді біріктіреді. Әрбір ереженің нәтижесі несиелік ұпайды көрсететін бір мәнге біріктіріледі.

Түйін сөздер: несиелік қабілеттілігі, шағын және орта бизнесті несиелендіру, бұлдыр логика, лингвистикалық айнымалы, логикалық ережелер, шешім қабылдау.

Қаржыландыру: Жұмысты ҚР ЖБФМ Ғылым комитеті, №АР0885497 гранты қолдады.

**У.Т. Махажанова^{1*}, Б. Тасуов², А.А. Муханова¹, А. Мухиядин¹,
Р.К. Жеткиншеков¹**

¹Евразийский национальный университет имени Л.Н. Гумилева,
Астана, Казахстан;

²Таразский региональный университет имени М.Х. Дулати, Тараз, Казахстан.
E-mail: *makhazhan.ut@gmail.com*

АЛГОРИТМ ОЦЕНКИ КРЕДИТОСПОСОБНОСТИ БИЗНЕСА НА ОСНОВЕ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ

Аннотация. Малый и средний бизнес играет значительную роль в экономическом развитии большинства стран. В условиях рыночной экономической системы одним из наиболее важных вопросов функционирования и развития малого бизнеса является поиск источников финансирования. Современное малое предприятие представляет собой открытую и динамически развивающуюся социально-экономическую и производственно-маркетинговую систему. Разнообразные экономические, финансовые, социальные и инвестиционные результаты, получаемые в результате производственно-хозяйственной деятельности, не могут быть оценены с помощью какого-либо одного показателя. Необходимо использование комплекса показателей, для достижения большей эффективности необходимо учитывать уровень развития малого предприятия, специфику деятельности, а также фактор неопределенности при получении того или иного финансового результата. В данной статье рассматриваются особенности кредитования малых предприятий. Чтобы определить целесообразность кредитования малого предприятия, рассматривался набор групп оценочных показателей: отраслевой и региональной специфики, деятельности малого предприятия и финансово-экономические показатели. Предложен метод оценки кредитоспособности предприятий малого и среднего бизнеса основанный на применении математического аппарата теории нечетких множеств. Метод нечеткого логического вывода использует нечеткие правила, которые позволяют встраивать в модель интуицию и опыт. Правила сочетают в себе показатели и коэффициенты. Результат каждого правила агрегируется в одно значение, которое представляет кредитный балл.

Ключевые слова: кредитоспособность, кредитование малого и среднего бизнеса, нечеткая логика, лингвистическая переменная, логические правила, принятие решений.

Финансирование: Работа поддержана Комитетом науки МВОН РК, грант № AP08855497.

**U. Makhazhanova^{1*}, B. Tassuov², A. Mukhanova¹, A. Mukhiyadin¹,
R. Zetkinshekov¹**

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

²Taraz Regional University named after M.Kh. Dulaty, Taraz, Kazakhstan.

E-mail: *makhazhan.ut@gmail.com*

AN ALGORITHM FOR ASSESSING THE CREDITWORTHINESS OF A BUSINESS BASED ON THE THEORY OF FUZZY SETS.

Abstract. Small and medium business plays a significant role in the economic development of most countries. In a market economic system, one of the most important issues for the functioning and development of small businesses is the search for sources of financing. A modern small business is an open and dynamically developing socio-economic, production and marketing system. A variety of economic, financial, social and investment results obtained as a result of production and economic activities cannot be assessed using any one indicator. It is necessary to use a set of indicators, in order to achieve greater efficiency, it is necessary to take into account the level of development of a small enterprise, the specifics of the activity, as well as the uncertainty factor when obtaining one or another financial result. This article discusses the features of lending to small businesses. To determine the feasibility of lending to a small business, a set of groups of estimated indicators was considered: industry and regional specifics, small business activities, and financial and economic indicators. A method for assessing the creditworthiness of small and medium-sized businesses based on the use of the mathematical apparatus of the theory of fuzzy sets is proposed. The fuzzy inference method uses fuzzy rules that allow intuition and experience to be built into the model. The rules combine indicators and coefficients. The result of each rule is aggregated into a single value that represents the credit score.

Key words: creditworthiness, lending to small and medium businesses, fuzzy logic, linguistic variable, logical rules, decision making.

Funding: The work was supported by the Scientific Committee of the MHES of the RK, grant No. AP08855497.

Кіріспе. Әлемдік дағдарыс жағдайында шағын және орта кәсіпкерлік саласындағы кәсіпкерлік қызмет жоғары тәуекелмен байланысты. Дегенмен, бұл секторды қаржыландыру мүмкіндіктері шектеулі болғандықтан несиелендіру ерекше маңызды. Шағын кәсіпкерлік секторын несиелендіру көбінесе кепілсіз несиелер болып табылады, бұл несие берушілерге қосымша

тәуекелдер жүктейді. Сондықтан банк үшін несие берушінің қажеттіліктеріне бейімделіп, оның мөлшерін ескере алатын және сонымен бірге икемді болып қала алатын қарыз алушының несие қабілеттілігінің өзіндік скорингтік моделінің болуы өте маңызды.

Қарыз алушының несиелік қабілеті мен несиелік тәуекелі компанияның ішкі сипаттамаларына және сыртқы жағдайларға (бар болса) негізделуі мүмкін. Несиелік тәуекелді талдау компанияның дефолт ықтималдығына негізделген, оны компанияның басқару стилі мен қаржылық, коммерциялық, саяси, іскерлік, салалық тәуекелдер деректері негізінде есептеуге болады. Несиелік қабілеттілігін талдау кезінде кәсіпорынның есеп беру коэффициенттері пайдаланылады, мысалы, ағымдағы өтімділік коэффициенті немесе пайдалылық коэффициенті. Банкрот өздерінің қалауларына байланысты әртүрлі коэффициенттер жиынтығын пайдаланады. Бірақ несиелеудің барлық түрлері үшін әмбебап коэффициенттер немесе әмбебап үлгі жоқ, деректерді тәжірибе мен парасаттылық негізінде талдау керек. Сондықтан баллдық үлгіні толтыратын кәсіпорындарды бағалау ережелерін құрастыру қажет.

Бағалау моделін жүзеге асыру әртүрлі болуы мүмкін. Дегенмен, екі бағытты бөліп көрсетуге болады: статистикалық әдістер және жасанды интеллект әдістері. Соңғылары көптеген салаларда бәсекелестеріне қарағанда жақсы жұмыс істейді және несиелік скоринг ерекшелік емес. Жасанды интеллектті жүзеге асыру тәсілдерінің бірі – бұлдыр логикалық ойларды көрсету үшін пайдаланылуы мүмкін нейрондық желі. Сондықтан, бұл мақалада біз қаржылық коэффициенттерді қамтитын сарапшылық ережелерге негізделген анық емес қорытынды жүйелеріне тоқталамыз. Бұл ережелер бұлдыр логика арқылы құрылады.

Әдебиеттерге шолу. Қолданыстағы қарыз алушы несиелік қабілеттілігін бағалау әдістері көп жағдайда статистикалық ақпараттарды талдауға негізделеді, шешілетін міндеттің ерекшелігіне қарай, қазіргі уақытта қажетті көлемде жетіспейді. Мысал ретінде – М. Керн, А. Хунгсок, А. Альтери, В.А. Чернов және т.б. зерттеу жұмыстарын айтуға болады.

Банкротты болжау әдістері болашақ қарыз алушы туралы тереңдетілген ақпарат алу үшін қолданылады. Кәсіпорынның банкрот болуын болжау дәлдігі жоғары, бірақ қысқа мерзімді. Қазіргі уақытта несиелік институттардың тәжірибелік қызметі қарыз алушының қаржы жағдайының төмендеу қаупін ескермейтін рейтингтік бағалау әдісін қолданады, ол несиелік тәуекел көлемін дұрыс есептемейді. Банкротты болжау және олардың негізінде қарыз алушыны бағалаудың рейтингтік моделін құрумен шетел және ресей ғалымдары айналысқан – Э. Альтман, А. Таффлер, У. Бивер, А. Фулмер, Дж. Чессар, Г.Г. Кадыкова, Г. Савицкой, К.С. Лаврушин және т.б.

Бұлдыр жиындар теориясына негізделген модельдерде баллдық жүйе бірқатар кезеңдерді қамтитын Бұлдыр қорытынды жасау процедурасына негізделген (Вазмага және б., 2014). Бұл жағдайда бастапқы факторлар мен нәтижелі атрибут мүшелік функцияларды қолдану арқылы нақтыланады және

логикалық ережелердің толық жүйесімен өзара байланысады (Nosratabadi және б., 2012).

Скорингтік модельдерге арналған нейрондық желілер қабаттарға топтастырылған нейрондардың жиынтығы болып табылады (Lai және б., 2006). Желінің кіріс деңгейі несие қабілеттілігін болжау үшін қолданылатын барлық сипаттамалардан тұрады, ал шығыс деңгейі клиенттің төлем қабілеттілігін көрсетеді (Pacelli және б., 2006). Скорингтік жүйелерді талдау листинглік зерттеулердің басым бөлігі дүниежүзілік ауқымда жүргізілгенін және акциялары қор нарықтарында тіркелген ірі трансұлттық компанияларға бағытталғанын көрсетті, бұл шағын бизнестің ерекшеліктерін ескермейді. Жоғарыда айтылғандарды қорытындылай келе, шағын және орта бизнесті кәсіпорындардың несие қабілеттілігін бағалау мәселесі өзекті, тәжірибелік маңызды, сонымен қатар жеткіліксіз әзірленген. Жүйе жұмыс жасайтын орта тұрақты емес, себебі жүйелердің жұмыс істеуінің сыртқы және ішкі жағдайларының өзгеруі несие беру шарттары мен несие қабілеттілігін сипаттайтын параметрлерді қайта қарау қажеттілігіне алып келеді. Осы себепті шағын және орта бизнесті кәсіпорындарды несиелендіру орындылығын талдауды, жоғары дәрежедегі белгісіздік жағдайында шешілетін есеп ретінде қарастыру қажет.

Материалдар мен әдістер. Бұлдыр логика классикалық логиканың жалпылауы болып табылады. 0 және 1-дің орнына $[0, 1]$ сегментіндегі кез-келген мәндер қолданылады, бұл ассоциативтілік дәрежесін арттыруға мүмкіндік береді.

Ресми түрде:

X әмбебап жиын болсын. X-дегі бұл бұлдырA жиыны реттелген жұптар жиыны ретінде анықталады.

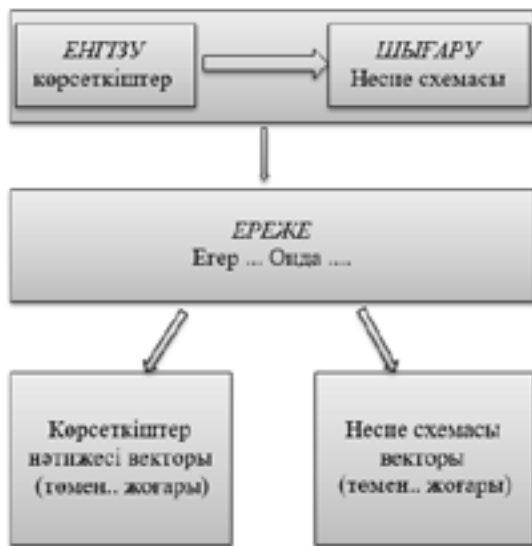
$$A\{x, \mu_A(x) | x \in X\},$$

$\mu_A(x)$ - A-дағы x мүшелік функциясы. $[0, 1]$ кесіндісінде X жиынының әрбір элементін байланыстырады.

Бірақ логикалық операцияларды кестеле көрсету мүмкін емес, логика шеңберінде олар функционалды түрде көрсетіледі:

$$\begin{aligned} a \wedge b &= \min(a, b), \\ a \vee b &= \max(a, b), \\ !a &= 1 - a. \end{aligned}$$

Бұлдыр логиканың мағынасы кіріс кеңістігін шығыс кеңістігімен салыстыру болып табылады. Мұның механизмі ережелер деп аталатын if-then операторларының тізімі. Ережелердің реті маңызды емес, сұраныстар қатар жүреді. Ережелерді түсіндіретін жүйені жасамас бұрын, біз пайдалануды жоспарлап отырған барлық терминдерді және оларды сипаттайтын тілдік айнымалыларды анықтауымыз керек. Жалпы схема:



Сурет 1 – ШОБ несие қабілеттілігін бағалау шешімі алгоритмі

Бұлдыр қорытынды - кіріс векторындағы мәндерді түсіндіретін және кейбір ережелер жиынтығына негізделген шығыс векторына мәндерді тағайындайтын әдіс. Бұлдыр қорытындылар жүйесін құру үшін шешім қабылдау ережелерін (яғни бұлдырережелер) құру және мүшелік функциялардың пішіні мен параметрлерін анықтау қажет. Бұл бастапқы кезеңде домен сарапшысы шешім қабылдау жүйесінде өзінің тәжірибесі мен білімін қалыптастырады. Бұлдыр логика бұл шешім ережелерін логикалық өрнекке түрлендіруге мүмкіндік береді. Домен сарапшылары шешім қабылдау ережелерін жасап, мүшелік функциялардың пішіні мен параметрлерін анықтағаннан кейін ережелерді түрлендіруді логика арқылы орындауға болады.

Нәтижелер. Кіріс деректерінің жиынтығын, кәсіпорынның бағалау көрсеткіштерінің жиынтығын анықтау.

Несиелік талдау табыстылық, өнімділік, өтімділік және төлем қабілеттілік салаларын қозғайды. Пайда коэффициенттері несиелік талдаушылар үшін маңызды, өйткені олар бизнестің қолайсыз бизнес жағдайларына жауап бере алатынын, жеткілікті түрде генерациялай алатынын көрсете алады. қарызға қызмет көрсетуден түскен пайда.

Тиімділік көрсеткіштері басшылықтың капитал, сату немесе жалпы бизнес көлемі арқылы пайда алу мүмкіндігін көрсетуі керек. Төлем қабілеттілігінің коэффициенттері несиелік талдау үшін маңызды, өйткені олар берешегі жоғары кәсіпорынның кеңейіп жатқанын көрсетеді. Төлем қабілеттілігінен басқа, несиелік талдаушылар әрқашан бизнестің өтімділігін тексереді, өйткені екі сала да оның табыстылығы үшін маңызды. Қаражаттың жеткіліксіз түсуі ұзақ мерзімді жоспарлаудың нашарлығынан кәсіпорынды банкроттыққа әкелуі мүмкін. Несие берушілер сондай-ақ компаниялардың

клиенттеріне несие беріліп жатқанына қызығушылық танытады және айналым капиталы ресурстарының қалай басқарылатынын тексеру үшін өнімділік коэффициенттерін пайдаланады. Егер бизнесте бизнес тәуекелі жоғары болса және пайда төмендеуге бейім болса, несие берушілер қарыздың төмен деңгейін көргісі келеді (Lebedeva және б., 2019).

Алдыңғы талқылаулар мен сарапшылардың білімдеріне сүйене отырып, кіріс параметрлерінің жиынтығынан кейбір коэффициенттерді бөліп көрсетейік (Абрамов және б., 2019).

Салалық және аймақтық ерекшелікті көрсеткіштер:

K_1 – Саланың даму динамикасы;

K_2 – Саланың даму болашағы;

K_3 – Мұндай өнімдерге (жұмысқа, қызметке) нарықтық (саланың) сұраныс;

K_4 – Аймақтық экономиканың даму динамикасы;

K_5 – Аймақтық экономиканың даму болашағы;

K_6 – Өнімнің (жұмыстың, қызметтердің) осы түріне нарықтық (аймақ) сұраныс.

Қаржылық-экономикалық көрсеткіштер және олардың ұсынылатын нормативті мәндері:

K_7 – Ағымдағы өтімділік коэффициенті, ұсынылатын мән ≥ 2 ;

K_8 – Қаржылық тәуелсіздік коэффициенті, ұсынылатын мән $\geq 50\%$;

K_9 – Меншікті айналым қаражатымен қорларды жабу үлесі, ұсынылатын мәнді $\beta \approx 60-80\%$ сарапшы анықтайды;

K_{10} – Борышты өтеу коэффициенті, ұсынылатын мән $\geq 1,5$ немесе 2;

K_{11} – Дебиторлық берешектің айналымдылығы, ұсынылатын мән ≤ 1 ;

K_{12} – Кредиторлық берешектің айналымдылығы, ұсынылатын мән ≥ 1 ;

K_{13} – Қорлардың айналымдылығы, ұсынылатын мән ≥ 6 ;

K_{14} – Меншікті және қарыз қаражатының арақатынас коэффициенті, ұсынылатын мән ≤ 1 ;

K_{15} – Өнімнің (сатудың) рентабельділік коэффициенті, ұсынылатын мән 40%.

Шағын және орта кәсіпорынның қызмет көрсеткіштері:

K_{16} – Қызметкерлердің кәсіби деңгейін бағалау;

K_{17} – Кәсіпорындағы моральды-психологиялық атмосфераны бағалау;

K_{18} – Кәсіпорынның нарықта болуының жеткіліктілігі;

K_{19} – Кәсіпорынның экономикалық саясаты;

K_{20} – Кәсіпорынның техникалық саясаты;

K_{21} – Кәсіпорынның кадрлық саясаты;

K_{22} – Қарыз алушының несие тарихы (жоқ).

Бағалау көрсеткіштерін $[0,1]$ аралығында көрсету.

Параметрлер жиыны лингвистикалық айнымалылар жиыны түрінде ұсынылады. Көп қажетті ақпараттар әдебиеттерде бар (Wang және б., 2015).

Жалпы жағдайда, кез-келген параметр кейбір интервалда нақты сандар мәнін қабылдайды деп есептейміз. Қарапайым жағдайда, әрбір параметрге

«осы көрсеткіштің қолайлылық дәрежесі» деп атауға болатын, бір лингвистикалық айнымалы сәйкестендіріледі. Нақты жағдайда, параметрге $[0,1]$ интервалы аралығын қабылдай алатындай нақты сан беру қажет. Осы процесс бірыңғайландыру процессі деп аталады. Егер бағалау параметрі мәні $\geq \alpha$, онда біз оны оң деп санаймыз (α мәнін сарапшы анықтайды), параметрді $[0,1]$ интервалы аралығын келтіру төмендегі шартпен жүзеге асады:

$$\mu_i(x) = \begin{cases} \frac{x}{2}, & \text{егер } x < \alpha, \\ 1, & \text{егер } x \geq \alpha. \end{cases}$$

Келтірілген функциялармен бірге «қолайлылық дәрежесі» композициясын алып, сәйкесінше лингвистикалық айнымалылар аламыз:

- параметр мәні төмен (L);
- параметр мәні орташадан төмен (LM);
- параметр мәні орташа (M);
- параметр мәні орташадан жоғары (HM);
- параметр мәні жоғары (H).

Осындай түрдегі лингвистикалық айнымалылар техникалық жүйелерде жиі қолданылады (Chourmouziadis және б., 2016). Оданда көбірек санды градациялар қолданылады: орташадан төмен, орташадан жоғары, нөлге жақын және т.б.

Шешім қабылдауы үшін бұлдыр жиындар теориясының математикалық аппаратын пайдалану.

Бұлдыр жиындар теориясы негізінде шағын бизнестің несие қабілеттілігін бағалауға арналған бұлдыр өндірістік жүйе, форманың функционалдық картасы

$$\begin{aligned} X &= \{x_i\} \rightarrow U = \{u_j\}, i = 1, \dots, n; j = 1, \dots, s, \\ U &= \{u_j\} \rightarrow Q, j = 1, \dots, s. \end{aligned}$$

мұндағы X – әсер етуші көрсеткіштердің векторы; U – жинақталған әсер етуші көрсеткіштер; Q - шағын бизнестің жалпы несие қабілеттілігі.

Қарыз алушыны несиелеудің орындылығын анықтау ұсынылатын $U = \{\mu_j\}$, $j = 1, \dots, s$ бағалау көрсеткіштерінің топтарының жиынтығы сапалық және сандық болып бөлінуі керек:

- салалық және аймақтық ерекшеліктердің көрсеткіштері;
- қаржылық-экономикалық көрсеткіштер;
- шағын кәсіпкерлік қызметінің көрсеткіштері.

Қарыз алушыға несие беру (несие беруден бас тарту) туралы шешім қабылдау A жиынының a_i критерийлері бойынша X жиынынан x_i көрсеткіштерінің картографиясын құру болып табылады. Нәтижесінде, критерийлер жиынтығын ескере отырып, әрбір x_i көрсеткіші үшін ол несиелік қабілеттіліктің өзіндік мәнін табу мүмкін болады.

Соңғы Q шығысын алу үшін көрсеткіштердің мәндерін қорытындылаймыз (Makhazhanova және б., 2020):

$$P_{LM}^1(x_1) \cap P_{LM}^2(x_2) \cap P_M^3(x_3) \cap P_M^4(x_4) \cap P_{HM}^5(x_5) \cap P_M^6(x_6) \cap P_{LM}^7(x_7) \cap P_M^8(x_8) \cap P_L^9(x_9) \cap P_L^{10}(x_{10}) \\ \cap P_M^{11}(x_{11}) \cap P_M^{12}(x_{12}) \cap P_L^{13}(x_{13}) \cap P_{LM}^{14}(x_{14}) \cap P_L^{15}(x_{15}) \cap P_{LM}^{16}(x_{16}) \cap P_M^{17}(x_{17}) \\ \cap P_{HM}^{18}(x_{18}) \cap P_M^{19}(x_{19}) \cap P_{LM}^{20}(x_{20}) \cap P_{LM}^{21}(x_{21}) \cap P_L^{22}(x_{22}) \rightarrow Q_M^1(y_1).$$

Q мәні несие беру немесе одан бас тарту туралы шешім қабылдауға мүмкіндік береді. Банкте қолжетімді несиелеу схемаларының нұсқаларын ескере отырып (S жиыны) және Q қорытынды рейтингінің мәні негізінде $S^*(S^* \in S)$ несиелеудің оңтайлы схемасы таңдалады.

Талқылау. Жұмыста шағын және орта бизнесті несиелеудің ерекшеліктері қарастырылған. Бұлдыр жиындар теориясының математикалық аппаратын пайдалану негізінде шағын кәсіпкерлік субъектілерінің несиелік қабілетін бағалау әдісі ұсынылған. Шағын бизнес – бұл ашық және қарқынды дамып келе жатқан әлеуметтік-экономикалық және өндірістік-маркетингтік жүйе. Өндірістік-шаруашылық қызмет нәтижесінде алынған әртүрлі экономикалық, қаржылық, әлеуметтік және инвестициялық нәтижелерді кез келген бір көрсеткішті пайдалана отырып бағалау мүмкін емес. Көрсеткіштер жүйесін (кешенін) пайдалану қажет. Бұл белгілі көрсеткіштердің ешқайсысы шағын кәсіпорындардың шаруашылық қызметінің сан алуан түпкілікті экономикалық, қаржылық және әлеуметтік нәтижелерін көрсете алмайтындығына байланысты. Сондықтан бұл жұмыста шағын кәсіпорынға тән салалық және аймақтық ерекшеліктердің көрсеткіштері, кәсіпорынның қызмет көрсеткіштері және қаржылық-экономикалық көрсеткіштер талданды. Жалпы жағдайда кез-келген параметр нақты сандардың белгілі бір интервалында мәндерді қабылдайды деп болжанады. Әрі қарай бастапқы көрсеткіштердің мәндері $[0,1]$ интервалында көрсетіледі. Бұл процесс деректерді унификациялау деп аталады.

Шешімдердің негізінде қабылданатын ережелер параметрлері бар логикалық формулалар түрінде қалыптасады. Несие қабілеттілігінің индексі деп аталатын ең жалпы түрде бір параметр болжанады, ол 0-ден 1-ге дейін өзгереді және табиғи түсіндірмесі бар. Несие қабілеттілігінің индексі жоғары кәсіпорындарға артықшылық беріледі. Несие берудің ең қолайлы схемасын таңдау критерийі ретінде ережелерді өңдеу нәтижесінде алынған бұлдыр жиынның мәнін пайдалануға болады (Zadeh және б., 1996). Сарапшылардың пікірлерін бағалау үшін интервал сандарын қолдануға болады. Сонымен қатар, интервалдық салмақтар шешімдер қабылдаудың әртүрлі мәселелерінде кеңінен қолданылады. Сенгупта мен Пал интервалдық сандарды салыстыру әдістерінің жан-жақты шолуын ұсынды (Sengupta және б., 2009). Екі интервал арасындағы артықшылықтың қарқындылығын сандық бағалау үшін ықтималдық өлшемдерін қолдану идеясы мүлдем жаңа емес (Zhang және б., 2017).

Басқару жүйелерінде бұлдыр логиканы қолданумен қатар оның негізінде сарапшылық жүйелердің жаңа буынын құру бойынша қарқынды жұмыстар жүргізілді. Атап өткендей (Коско және б., 1993), бұлдыр сараптамалық жүйелер өздерінің негізгі артықшылығына – нақты дүние жағдайларына жақсырақ бейімделгеніне қоса, дәстүрліге қарағанда тағы екі артықшылығы бар.

Біріншіден, олар деп аталатын нәрселерден босатылады. қорытындыларды құрудағы «циклдік құлыптар». Екіншіден, бұлыңғыр ережелердің әртүрлі негіздерін оңай біріктіруге болады, бұл әдеттегі сараптамалық жүйелерде сирек мүмкін.

Қорытынды. Қарыз алушы шағын кәсіпорын несиелендірілу қабілеттілігін бағалауға арналған, сапалақ және сандық көрсеткіштері бар бұлдыр базаларға негізделіп әзірленген әдісті қолдану, төмендегідей мүмкіндіктер береді:

- шешім қабылдау процессінде несиелендірілу сарапшысының қатысуы есебінен несиелендірілу қабілеттілігіне жүргізілген талдау сапасын сақтайды, сонымен қатар бұлдыр жиындар математикалық аппаратты қолдану есебінен сараптамалық бағалау дұрыс жұмыс жасайды;

- шағын кәсіпорынды несиелендіру кезінде коммерциялық банк шығындарын азайтады және несиелендіру беру бойынша шешім қабылдау процессіне теріс әсер ететін факторларды бейтараптандырады;

- кәсіпорынның несиелендірілу қабілеттілігін бағалауды динамикалық процесс ретінде жүзеге асырады, яғни несиелендірудің барлық мерзімі ішінде, бұл банкке қарыз алушы күйіне бақылау жүйесін ұйымдастыруға мүмкіндік береді, сонымен қатар несиелендіру қаупі басталу кезеңінде хабар алады.

- несиелендіру сарапшысы қалауына байланысты бағалау көрсеткіштері жиынын ауыстыру, кәсіпорын қызметі ерекшелігі саласына байланысты немесе қарыз алушы ұсынған қаржы ақпаратына сәйкес, стандартты қаржы есебі ретінде қолдануға болатын, қысқартылған салық жүйесі деректері;

- бірыңғай автоматтандырылған деректер базасын қалыптастыру, сала және аймақ бойынша даму динамикасы бағасы көрсеткіштерін, банк қайта сапалық көрсеткіштерді ұқсас кәсіпорындар үшін есептемес үшін;

- несиелендіру көлемі ұлғаяды.

Ұсынылған ғылыми тәсіл әлеуметтік-экономикалық қызметтің әртүрлі салаларында сараптамалық шешімдерді қолдау жүйелерін құру, процестерді бақылау және кәсіпорынның қаржылық-шаруашылық қызметін талдау үшін негіз ретінде пайдаланылуы мүмкін.

Information about the authors:

U. Makhazhanova – PhD, Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: makhazhan.ut@gmail.com, <https://orcid.org/0000-0002-5528-8000>;

B. Tassuov – Candidate of technical sciences, Department of Physics and Information Technology, Taraz Regional University named after M.Kh. Dulaty, Taraz, Kazakhstan, E-mail: bolat_ktn@mail.ru, <https://orcid.org/0000-0002-2000-6720>;

A. Mukhanova – PhD, Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana,

Kazakhstan, E-mail: ayagoz198302@mail.ru, <https://orcid.org/0000-0003-3987-0938>;

A. Mukhiyadin – Doctoral student of the Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: amukhiyadin@gmail.com, <https://orcid.org/0000-0001-5576-7733>;

R. Zhetkinshekov – Master degree student of the Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: zhrk2407@gmail.com.

ӘДЕБИЕТТЕР:

Абрамов Е.П., Махажанова У.Т., Мурзин Ф.А. Принятие решений в области кредитования на основе размытой логики Заде. 12-я международная Ершовская конференция по информатике (PSI'19), Россия, Новосибирск, 2019, 20-25.

Sengupta A., Pal T.K. On comparing interval numbers: a study on existing ideas. *Fuzzy Preference Ordering of Interval Numbers in Decision Problems*, Springer Berlin Heidelberg, 2009, 25-37. https://doi.org/10.1007/978-3-540-89915-0_2.

Bazmara A., Donighi S.S. Bank Customer Credit Scoring by Using Fuzzy Expert System. *International Journal of Intelligent Systems and Applications*, 2014, 11, 29–35 <https://doi.org/10.5815/ijisa.2014.11.04>.

Kosko, Bart. *Fuzzy thinking* / Hyperion, 1993.

Chourmouziadis K., Chatzoglou P.D. An intelligent short term stock trading fuzzy system for assisting investors in portfolio management. *Expert Systems with Applications*, 2016, 43, 298-311. <https://doi.org/10.1016/j.eswa.2015.07.063>.

Lai K.K., Yu L., Zhou L.G., Wang S.Y. Neural Network Metalearning for Credit Scoring. *International Conference on Intelligent Computing*, 2006, 403-408. https://doi.org/10.1007/11816157_47.

Lebedeva M.E. Fuzzy logic in economics – the formation of a new direction. *Ideas and Ideals*, 2019, 11(1), 197–212. <https://doi.org/10.17212/2075-0862-2019-11.1.1-197-212>.

Makhazhanova U.T., Murzin F.A., Mukhanova A.A., Abramov E.P. Fuzzy logic of Zadeh and decision-making in the field of loan. *Journal of theoretical and applied Information Technology*, 2020, 98 (06), 1076-1086.

Nosratabadi H.E., Nadali A., Pourdarab S. Credit Assessment of Bank Customers by a Fuzzy Expert System Based on Rules Extracted from Association Rules. *International Journal of Machine Learning and Computing*, 2012, 2(5), 662–666.

Wang Y.-J. Ranking triangle and trapezoidal fuzzy numbers based on the relative preference relation. *Applied Mathematical Modelling*, 2015, 39(2), 586-599 <https://doi.org/10.1016/j.apm.2014.06.011>.

Pacelli V., Azzollini M. An Artificial Neural Network Approach for Credit Risk Management. *Journal of Intelligent Learning Systems and Applications*, 2011, 3, 103–112. <https://doi.org/10.4236/jilsa.2011.32012>.

Zadeh L.A. Shadows of fuzzy sets. *Advances in Fuzzy Systems – Applications and Theory Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems*, 1996, 51-59. https://doi.org/10.1142/9789814261302_0004.

Zhang Z. Logarithmic least squares approaches to deriving interval weights, rectifying inconsistency and estimating missing values for interval multiplicative preference relations. *Soft Computing*, 2017, 21, 3993-4004. <https://doi.org/10.1007/s00500-016-2049-6>.

REFERENCES:

Abramov E.P., Makhazhanova U.T., Murzin F.A. Decision-making in the field of lending based on the blurred logic of the loan. 12th International Ershov Conference on Computer Science (PSI'19), Russia, Novosibirsk, 2019, 20-25.

Sengupta A., Pal T.K. On comparing interval numbers: a study on existing ideas. Fuzzy Preference Ordering of Interval Numbers in Decision Problems, Springer Berlin Heidelberg, 2009, 25-37. https://doi.org/10.1007/978-3-540-89915-0_2.

Bazmara A., Donighi S.S. Bank Customer Credit Scoring by Using Fuzzy Expert System. International Journal of Intelligent Systems and Applications, 2014, 11, 29–35 <https://doi.org/10.5815/ijisa.2014.11.04>.

Kosko, Bart. Fuzzy thinking / Hyperion, 1993.

Chourmouziadis K., Chatzoglou P.D. An intelligent short term stock trading fuzzy system for assisting investors in portfolio management. Expert Systems with Applications, 2016, 43, 298-311. <https://doi.org/10.1016/j.eswa.2015.07.063>.

Lai K.K., Yu L., Zhou L.G., Wang, S.Y. Neural Network Metalearning for Credit Scoring. International Conference on Intelligent Computing, 2006, 403-408. https://doi.org/10.1007/11816157_47.

Lebedeva M.E. Fuzzy logic in economics – the formation of a new direction. Ideas and Ideals, 2019, 11(1), 197–212. <https://doi.org/10.17212/2075-0862-2019-11.1.1-197-212>.

Makhazhanova U.T., Murzin F.A., Mukhanova A.A., Abramov E.P. Fuzzy logic of Zadeh and decision-making in the field of loan. Journal of theoretical and applied Information Technology, 2020, 98 (06), 1076-1086.

Nosratabadi H.E., Nadali A., Pourdarab S. Credit Assessment of Bank Customers by a Fuzzy Expert System Based on Rules Extracted from Association Rules. International Journal of Machine Learning and Computing, 2012, 2(5), 662–666.

Wang Y.-J. Ranking triangle and trapezoidal fuzzy numbers based on the relative preference relation. Applied Mathematical Modelling, 2015, 39(2), 586-599 <https://doi.org/10.1016/j.apm.2014.06.011>.

Pacelli V., Azzollini M. An Artificial Neural Network Approach for Credit Risk Management. Journal of Intelligent Learning Systems and Applications, 2011, 3, 103–112. <https://doi.org/10.4236/jilsa.2011.32012>.

Zadeh L.A. Shadows of fuzzy sets. Advances in Fuzzy Systems – Applications and Theory Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems, 1996, 51-59. https://doi.org/10.1142/9789814261302_0004.

Zhang Z. Logarithmic least squares approaches to deriving interval weights, rectifying inconsistency and estimating missing values for interval multiplicative preference relations. Soft Computing, 2017, 21, 3993-4004. <https://doi.org/10.1007/s00500-016-2049-6>.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 4, Number 344 (2022), 93-105
<https://doi.org/10.32014/2022.2518-1726.159>

UDC 004.574.58

**R.M. Moldasheva^{1*}, A.A. Ismailova¹, A.K. Zhamangara², A.M. Zadagali³,
G.B. Turmukhanova⁴**

¹S. Seifullin Kazakh Agrotechnical University, Astana, Kazakhstan;

²Astana botanical garden is a branch RSE REM «Institute of botany and phytointroduction», CFW MEGNR RK, Astana, Kazakhstan;

³L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

⁴Atyrau University named after H.Dosmukhamedov, Atyrau, Kazakhstan.

E-mail: raushan85_07@mail.ru

REQUIREMENTS TO DEVELOPMENT OF IAS FOR RESEARCH OF AQUEOUS ECOSYSTEMS

Abstract. The purpose of the article is to describe the concept of developing an information-analytical system for monitoring the quality of water resources. The main goal is to develop methods and techniques for regulating the temperature and hydrochemical regime of water in water bodies under the influence of anthropogenic and natural factors.

Information parameters of monitoring were selected and substantiated, structural scheme of the information-analytical system was developed and described. The proposed concept of the system with application of algorithms based on fuzzy logic allows to achieve favorable conditions in comparison with non-automated open reservoirs. The aim of the work is to determine the structure of information and analytical support of water quality monitoring in frame of the task of studying information parameters of temperature and hydrochemical state of water bodies.

In order to substantiate the concept of creating information-analytical support for water quality monitoring it is necessary to define the information parameters within the norm and the ranges of their changes. Based on analysis of scientific literature a number of information parameters were defined: water temperature (15-23°C), the level of dissolved oxygen (standard value), salts dissolved in water. An important indicator of a suitable environment is the level of oxygen dissolved in water. Lack of oxygen leads to a metabolic disorder, while excess leads to formation of gas bubbles in the vessels. The amount of salts dissolved in the water affects the level of osmotic pressure and is the basis for development of fish food organisms.

An information-analytical system has been developed that allows collecting and systematizing of information about the state of water resources and changes thereof, on the sources and factors of impact, on the introduction of anthropogenic burden thereon. An algorithm for assessing water quality based on associative indicators has been developed allowing a rapid assessment and making of timely optimal decisions in the development of environmental protection action. In general this article presents the concept of system organization: defining criteria for assessing "current state" of water bodies; identifying and preventing the development of anomalies with a high level of reliability and accuracy.

There fore further development and implementation of the system using algorithms based on fuzzy logic allows to achieve more favorable conditions in comparison with non-automated open water bodies.

Key words: mathematical model, information-analytical system, aquatic ecosystem, database, phytoplankton, biomass.

**Р.Н. Молдашева^{1*}, А.А. Исмаилова¹, А.К. Жамангара², А.М. Задағали³,
Г.Б. Турмуханова⁴**

¹С. Сейфуллин атындағы Қазақ агротехникалық университеті,
Астана, Қазақстан;

²Астана ботаникалық бағы – ҚР ЭГТРМ ОШЖДК «Ботаника және
фитоинтродукция институты» ШЖҚ РМК филиалы, Астана, Қазақстан;

³Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

⁴Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан.

E-mail: raushan85_07@mail.ru

СУ ЭКО ЖҮЙЕЛЕРІН ЗЕРТТЕУДЕ АТЖ ӘЗІРЛЕУГЕ ҚОЙЫЛАТЫН ТАЛАПТАР

Аннотация. Мақала су ресурстары сапасының мониторингінің ақпараттық-талдау жүйесін құру тұжырымдамасының сипаттамасына арналған. Техногендік және табиғи факторлардың әсеріне ұшыраған су айдындарындағы судың температуралық және гидрохимиялық режимін бақылау әдістері мен тәсілдерін әзірлеу негізгі мақсат болып табылды. Мониторингтің ақпараттық параметрлері таңдалды және негізделді, ақпараттық-талдау жүйесінің құрылымдық схемасы әзірленіп сипатталды. Анық емес логикаға негізделген алгоритмдерді қолданатын жүйенің ұсынылған тұжырымдамасы автоматтандырылмаған ашық су объектілерімен салыстырғанда қолайлы жағдайларды сақтауға қол жеткізуге мүмкіндік береді.

Жұмыстың мақсаты су айдындарындағы судың температуралық және гидрохимиялық режимінің ақпараттық параметрлерін зерттеу міндеттеріндегі су ресурстары сапасының мониторингін ақпараттық-талдамалық қамта-

масыз ету құрылымын айқындау болып табылады. Су ресурстары сапасының мониторингін ақпараттық-талдамалық қамтамасыз етуді құру тұжырымдамасын қолдау үшін норма шегінде ақпараттық параметрлерді және олардың өзгеру диапазондарын айқындауды талап етеді. Ғылыми әдебиеттерді талдау негізінде бірқатар ақпараттық параметрлер анықталды: судың температурасы (15-23⁰C), ерітілген оттегінің деңгейі (норма), суда ерітілген тұздар.

Қолайлы ортаның маңызды көрсеткіші суда ерітілген оттегінің деңгейі болып табылады. Оттегінің жетіспеушілігі метаболизмнің бұзылуына, артық болуы қан тамырларында газ көпіршіктерінің пайда болуына әкеледі. Суда еріген тұздардың мөлшері осмотикалық қысым деңгейіне әсер етеді және балық үшін қоректік организмдердің дамуына негіз болады. Ақпараттық-аналитикалық жүйе әзірленді, ол су ресурстарының жай-күйі және олардың өзгерістері туралы, әсер ету көздері мен факторлары туралы, оларға техногендік жүктемелердің жол берілуі туралы ақпаратты жинақтауға және жүйеге келтіруге мүмкіндік береді. Ассоциативті көрсеткіштер бойынша су сапасын бағалау алгоритмі жасалды, бұл экологиялық шараларды әзірлеу кезінде тез бағалауға және оңтайлы шешімдерді уақтылы қабылдауға мүмкіндік береді. Жалпы осы мақалада ұсынылған жүйені ұйымдастыру тұжырымдамасы: су объектілерінің «ағымдағы жай-күйін» бағалау өлшемдерін анықтау; сенімділік пен дәлдіктің жоғары деңгейімен нормадан ауытқулардың дамуын анықтау және алдын алу.

Демек, анық емес логикаға негізделген алгоритмдерді қолдана отырып, жүйені одан әрі дамыту және іске асыру автоматтандырылмаған ашық су объектілерімен салыстырғанда қолайлы жағдайларды сақтау арқылы қол жеткізуге мүмкіндік береді.

Түйін сөздер: математикалық модель, ақпараттық-аналитикалық жүйе, су экожүйесі, деректер қоры, фитопланктон, биомасса.

**Р.Н. Молдашева^{1*}, А.А. Исмаилова¹, А.К. Жамангара², А.М. Задағали³,
Г.Б. Турмуханова⁴**

¹Казахский агротехнический университет им. С. Сейфуллина,
Астана, Казахстан;

²Астанинский ботанический сад – филиал РГП на ПХВ «Институт ботаники
и фитоинтродукции» КЛХЖМ МЭГПР РК, Астана, Казахстан;

³Евразийский национальный университет им. Л.Н. Гумилева,
Астана, Казахстан;

⁴Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан.

E-mail: raushan85_07@mail.ru

ТРЕБОВАНИЯ К РАЗРАБОТКЕ ИАС-ИССЛЕДОВАНИЙ ВОДНЫХ ЭКОСИСТЕМ

Аннотация. Статья посвящена описанию концепции создания информационно-аналитической системы мониторинга качества водных ресурсов. Основная цель заключалась в разработке методов и методов регулирования температурного и гидрохимического режима воды в водоемах, находящихся под воздействием техногенных и природных факторов. Выбраны и обоснованы информационные параметры мониторинга, разработана и описана структурная схема информационно-аналитической системы. Предложенная концепция системы с использованием алгоритмов на основе нечеткой логики позволяет добиться благоприятных условий по сравнению с неавтоматизированными открытыми водоемами. Цель работы – определить структуру информационно-аналитического обеспечения мониторинга качества водных ресурсов в задаче исследования информационных параметров температурного и гидрохимического режима воды водных объектов. Для обоснования концепции создания информационно-аналитического обеспечения мониторинга качества водных ресурсов необходимо определить информационные параметры в пределах нормы и диапазоны их изменения. На основании анализа научной литературы определен ряд информационных параметров: температура воды (15-23°C), уровень растворенного кислорода (норма), растворенные в воде соли. Важным показателем подходящей среды является уровень растворенного кислорода в воде. Недостаток кислорода приводит к нарушению обмена веществ, а избыток – к образованию пузырьков газа в сосудах. Количество растворенных в воде солей влияет на уровень осмотического давления и является основой развития пищевых организмов рыб. Разработана информационно-аналитическая система, позволяющая собирать и систематизировать информацию о состоянии водных ресурсов и их изменениях, об источниках и факторах воздействия, о введении на них техногенных нагрузок. Создан алгоритм оценки качества воды на основе ассоциативных показателей, позволяющий проводить быструю

оценку и своевременно принимать оптимальные решения при разработке природоохранных мероприятий. В целом в данной статье представлена концепция организации системы: определение критериев оценки «текущего состояния» водных объектов; выявлять и предотвращать развитие аномалий с высоким уровнем достоверности и точности. Поэтому дальнейшее развитие и внедрение системы с использованием алгоритмов на основе нечеткой логики позволяет добиться более выгодных условий по сравнению с неавтоматизированными открытыми водоемами.

Ключевые слова: математическая модель, информационно-аналитическая система, водная экосистема, база данных, фитопланктон, биомасса.

Introduction. Natural resources of a society including water ones largely determine the degree and possibilities of its sustainable development. Evolution of life style, population growth, development of production and other types of human activity are related to consumption of various resources including water ones. The main challenges of the Republic of Kazakhstan water sector include a high level of anthropogenic load on water bodies: pollution by various chemicals, violation of the natural regime during construction of hydraulic-technical structures, irretrievable water withdrawals.

Figure 1 shows the water ecosystem and sources of contamination, and spells out the consequences of contamination.

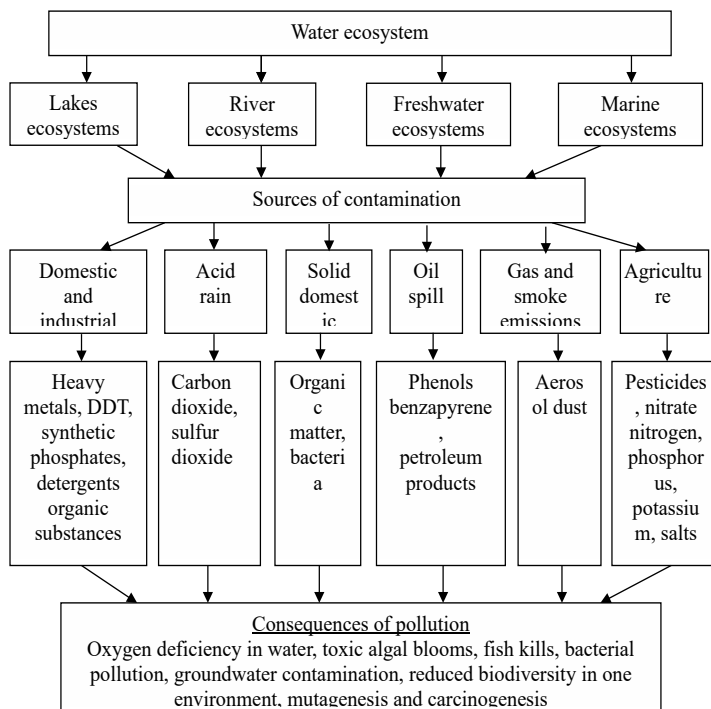


Figure 1 - Water ecosystem

Water pollution is a large-scale environmental problem, however modern methods of treatment do not solve it. It can lead to serious environmental consequences, since no living creature can survive without water. In order to solve the problem it is necessary to determine the sources of pollution and existing approaches to handle them.

The issue of pollution of surface natural waters has grown to the scale of the planet. Mankind is trying to solve the problem in different ways (<https://cleanbin.ru/problems/water-pollution>).

Research materials and methods. Innovations, technological solutions, tax incentives and huge fines are the methods that can together solve the problem of water pollution. Otherwise, already in the middle of the XXI century more than 80% of the population will experience an acute shortage of clean water sources. Today we are already talking about 45-50% of people, which figure indicates a threatening trend (<https://vyvoz.org/blog/zagrzaznenie-vody/>).

The Government also performs continuous monitoring. The Water Code of the Republic of Kazakhstan states that the Government monitoring of water bodies is a system of regular observations of hydrological, hydrogeological, hydrogeochemical, sanitary-chemical, microbiological, parasitological, radiological and toxicological indicators of state thereof, collection, processing and transmission of obtained information with the purpose of timely identification of negative processes, assessing and forecasting of development thereof, development of recommendations on prevention of harmful consequences and determination of the efficiency level of the water management measures.

The concept of monitoring for complex objects of artificial origin defines monitoring as the process of tracking the state of an object and comparing its current state with reference samples (levels or points).

Monitoring of a natural object is understood as a process of collecting, storing, transforming experimental information about the object of monitoring and development of recommendations in support of some or other decisions that are optimal in terms of the criteria adopted.

Problems of the monitoring system in Kazakhstan:

- poorly funded system;
- underdeveloped monitoring network;
- information collection is insufficient and non-systematic, outdated methods and equipment are applied;
- poor technical equipment of the monitoring network and insufficient application of modern methods;
- inconsistency in operation of monitoring services of different agencies and uncoordinated location of the monitoring network (Source: MEP, UNDP IWRM project (2005)).

Management in the field of environmental protection is a subordinate executive and administrative activity of state executive authorities and local administrations,

the main purpose of which is to organize preservation of the environment, recovery thereof and ensuring environmental safety (Krassov, 2014: 624).

One of the most important problems in the development of the information model of the object under environmental monitoring is the choice of information sources:

- cartography including data on the geographical location of regions and the functional use of territories;
- information about energy production and energy consumption structure in the regions;
- sources of anthropogenic pollution;
- data arriving from stationary posts of environmental monitoring, hydrometeorological changes;
- the results of the sampling analysis of the medium;
- aerospace sounding;
- medical-biological and social research, etc.

It should be noted that there is limited information on the actual quality of water in water bodies and the location and nature of the main pollution sources: the quality of surface water throughout the country is monitored by only 215 monitoring stations of Kazgidromet. There is no national register of the main pollution sources and the main pollutants available (Naidenko et al, 2003: 186).

The situation is also aggravated by the lack of data, insufficient quality monitoring and implementation of measures to improve efficiency as well as low level compliance with standards and requirements. There is no integrated national database containing critical information on water balance and consumption in different sectors, there is no integrated electronic database on the state of water bodies (Schreider, 1994: 72; Tsvetkov, 1998: 288).

Thus in terms of information tasks of environmental quality management the main problems are:

- there is no prediction or complicated prediction of the state of the aquatic environmental systems depending on the actions of managing authorities and the state of objects under monitoring;
- the results of assessment or forecast do not reach those for whom they are intended, or are presented in a way that the addressee does not perceive them.

Inefficient operation of traditional systems of receiving, processing and transmitting the information leads to violations in the systems of decision-making and control actions. For successful management of lakes resources in Kazakhstan it is necessary to develop an information-logical model of the subject area, which reflects generalized characteristics of aquatic ecosystems state and it is necessary to be able to promptly and in a visual form obtain detailed information about controlled objects necessary for decision-making.

The information-analytical system of research of aquatic ecosystems being

developed is suitable for use in applied environmental protection activities. Users of the developed information-analytical system might be specialists of the chosen industry and might have no mathematical or computer education (ST of RoK 34.005-2002. Information technology. Basic terms and definitions. - Introduced on 2004.01.01. - TC #34 on standardization "Information technologies". - Astana, 2002: 36). IAS refers to a special class of information systems, which is used for analytical processing of hydrobiological and hydrochemical data.

Results. As of today the class of information-analytical systems is one of the most demanded in the software market, it is related to the process of automation of information activity of aquatic ecosystems research.

The current system of environmental monitoring carried out by both scientific institutions (Republic of Kazakhstan Institute of Geography, Kazakh Scientific-Research Institute of Water Management, Kazakh Scientific-Research Institute of Fishery, information and analytical centers of environmental protection) and republican supervisory bodies is low effective not only because of insufficient technical equipment, but also largely due to ignoring of modern data management methods and integrated mathematical processing of multi-dimensional observations. The richest material on the hydrochemistry of natural water systems (accumulated during decades by the regional services of Kazhydromet) and the data of hydrobiological observations (accumulated by the Committee for Environmental Regulation and Control and the Kazakhstan Agency of Applied Ecology) remains undemanded and is lost every year. It is obvious that in addition to traditional low-informative summaries of the proportion of indicators exceeding MACs, these data could be successfully used to build both local models of seasonal and long-term dynamics of water bodies and generalized models of rational environmental and economic development of territorial complexes (Shitikov et al, 2003: 463).

The following requirements are common for the IAS of aquatic ecosystems research in general:

In terms of architectural construction:

1) have a three-link "client-server" architecture: client, application server and database server;

2) user access to functional services should be carried out by "thin client" technology on the basis of web-browser through an automated workstation;

3) The browsers of at least 7.0 Internet Explorer family must be supported;

In terms of indicators of information interaction:

4) have a centralized data repository that allows to maintain unified directories and perform centralized control of directories replenishment;

5) be able to perform operations of arranging information interaction and exchange.

In terms of indicators of user information needs satisfaction:

6) ensuring completeness, sufficiency, relevance and reliability of information.

In terms of interaction with the system:

7) multi-user mode support;

8) definition of user rights to access services, functions and data depending on the user's role in the system;

9) the possibility of scaling the performance and scope of processed information without modifying its software by upgrading the hardware complex used;

10) providing the ability to store the data historically.

The following shall be taken into account in the System:

- The prospect of integrating the System with other information systems of the Ministry of Environment and Water Resources;

- the level of readiness of potential users to use the System and the possibility of improving the general and professional culture of information interaction;

- real and prospective needs of potential users for information resources and services of the System;

- ensuring the possibility of upgrading under any organizational and functional changes in the Ministry of Environment and Water Resources of the Republic of Kazakhstan subject to preservation and possibility of further use of all necessary information.

The basic principles of the System are:

- the use of industrial software and hardware;

- the architecture of the System is built according to the principle of client-server application. The communication channels of Internet providers or channels in local networks (Intranet) shall be used as data transmission method between the client and the server (Standard of the RoK 34.019-2002. Information technology. Processes of software life cycle. - Introduced on 2004.01.01. - TC № 34 on standardization "Information technology". - Astana, 2002: 33).

To work comfortably with the System the bandwidth of communication channels shall be no more than:

Internet - 54 Kbps per client (the speed can be higher when using dedicated communication channels);

Intranet - 100 Mbps.

Figure 2 shows a possible scheme of connecting clients to the database server.

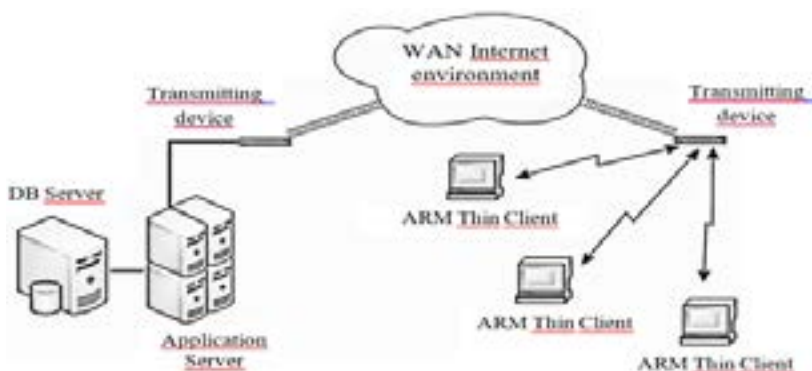


Figure 2 - System hardware architecture

The hardware requirements shown in Table 1 are the minimum recommended configuration, at which the system will be operable.

Table 1 - Minimum recommended configuration of hardware

Equipment	Technical characteristics
1	2
Web server	
Processor	4 quad-core 2.4 GHz
RAM	32 GB
Discs	2 drives 146 GB SAS 10K (RAID 1)
OS	Windows Server 2008 R2 Standard x64 and above or Windows Server 2012 R2 Standard x64 and above or Ubuntu Server 14.04 LTS and above
Database server	
Processor	4 quad-core 2.4 GHz
RAM	32 GB
Disks for the system	2 50GB SAS 10K drives (RAID 1)
Data storage drives	2 1TB SAS 10K drives (RAID 1)
OS	Selected for the web server
Test server	
Processor	2 quad-core 2.4 GHz
RAM	16 GB
Disks for the system	1 146 GB SAS 10K drive
Data storage drives	1TB SAS 10K drive
OS	Selected for the web server
Client computer (for thin client automated workstation)	
Processor	2 GHz
RAM	2 GB
Free space on disk	10 GB
Monitor	19" and above Resolution 1600x900 and higher
Keyboard	
Mouse	
OS	Windows 7 and later

The application server provides data retrieval from the database, calculation and formatting of data in accordance with the requirements of a particular report.

The Application server can be technically combined on the same computer as the Database server. But such a combination may reduce the overall performance of the System. Testing of the System was performed on the combined Windows 2007 server and showed good practical results.

The thin client automated workstation is a Microsoft Internet Explorer 7 application.

The proposed technology is well-proven in local and corporate networks, but can also work in the open WWW-space. But for security reasons it is not recommended to open WWW-access.

All automated workstations (AWS) are workstations.

The system requires mandatory registration or login by user password. To search, select, and view information a system of typical queries to the database with a specially developed system of menus and dialogs is used (GOST R 53622-2009. Information technologies. Information-computer systems. Stages and life cycle phases, types and completeness of documents. - Introduced on 2002-01-01. - Moscow: Publishing house of standards, 2001: 27).

The main functionalities of the information subsystem:

- input, editing and processing of information;
- making of different queries to databases, search for objects in spreadsheets;
- quick retrieval of required information (by year, by type, by name);
- calculation of the maximum, average, total value of the selected search parameter;
- calculation of the time interval and values exceeding the MAC of harmful substances for fishery bodies of water by certain parameters;
- output of the results in various forms including printing.

Discussion. Observation and facts are important aspects of the environmental systems analysis. Along with these a mathematical model is equally important in understanding environmental and ecological preservation problems. The software implementation of mathematical models of ecological systems on personal computers - computer models of ecological systems gives the opportunity to conduct computational experiments to describe ecological processes.

Observation and measurement of the characteristics of ecological systems are obtained as a result of environmental monitoring and laboratory studies, which are the initial data for development of mathematical models. Figure 3 shows the procedure for developing a mathematical model and computer aided simulation of the ecological system.

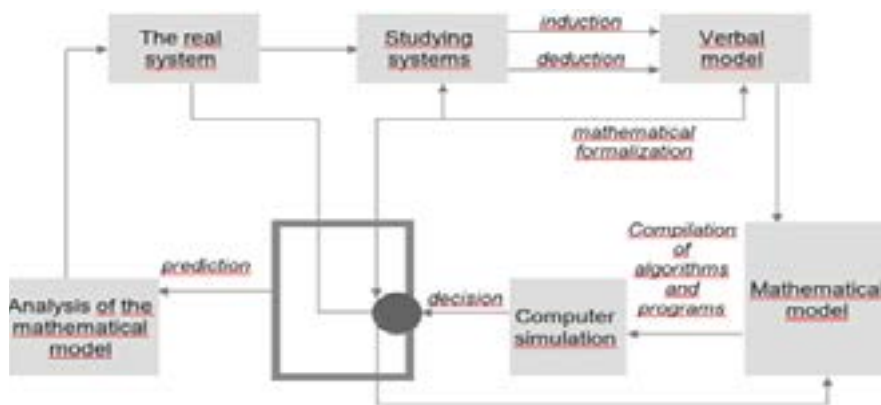


Figure 3 - Procedure for developing a mathematical model and computer aided modeling of the ecosystem

To conduct computer aided modeling and computational experiment it is necessary to create algorithms and sets of programs. This can be done on the basis of a changing mathematical model. The results obtained in the computational experiment and computer aided simulation of ecological systems are used in comparison with the knowledge of ecological systems and the environment and with the results of observation.

Special statistical methods are used to estimate the parameters of mathematical models when identifying ecosystems. They serve to interpret the available quantitative information about the relations within the ecosystem.

Algorithmization and programming of systems of equations of mathematical model for computer aided modeling of ecosystems can be carried out with application of various programming languages and by means of various special application software packages. After choosing the programming language for computer aided modeling of ecosystems it is necessary to perform the following procedures (Meshalkin et al, 2020: 357):

- Creating a functional block diagram of the model indicating the variables used for each block of the block diagram and the calculation steps.

- Writing a program in the selected language for each individual block of the computer model.

- Program debugging and correction of syntactic and logical errors made during programming.

- Checking the correctness of the computer aided model.

Conclusion. The main task of the computer aided modeling of ecosystems is to simulate processes and phenomena in an ecosystem or specific calculations under given conditions. The results of computer aided modeling of ecosystems can be used as a benchmark for assessing changes in ecosystem behavior when environmental conditions change.

Information about the authors:

Moldasheva Raushan Nurkozhaevna – Doctoral student of the Department of «Information Systems» S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; Personal phone number: 8-701-9940336, *raushan85_07@mail.ru*, 0000-0002-4570-0487;

Ismailova Aisulu Abzhapparovna – PhD, Associate Professor of the Department «Information Systems» S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; Personal phone number: 8-701-4606049, *a.ismailova@mail.ru*, 0000-0002-8958-1846;

Zhamangara Aizhan Kashagankyzy – Candidate of Biological Sciences, Deputy director for science of the Astana botanical garden branch of the RSE REM «Institute of botany and phytointroduction» CFW MEGNR RK, Nur-Sultan, Kazakhstan; Personal phone number: 8-707-8485060, *Kashagankizi@mail.ru*, 0000-0002-2348-1711;

Zadagali Aizhan Meirangalikyzy – Doctoral Student of the Department of «General Biology and Genomics» L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan; Personal phone number: 8-701-6677064, z.a.aizhan1993@gmail.com, 0000-0002-2537-3538;

Turmukhanova Gulnur Boranbayeva – Senior lecturer of the Department of «Software Engineering» Atyrau University named after H. Dosmukhamedov, Atyrau, Kazakhstan; Personal phone number: 8-778-662-26-13, turmukhanovag@mail.ru, 0000-0002-8912-2746.

REFERENCES

GOST R 53622-2009. Information technologies. Information-computer systems. Stages and life cycle phases, types and completeness of documents. - Introduced on 2002-01-01. - Moscow: Publishing house of standards, 2001.-27 p.

Krassov O.I. Environmental Law: textbook. - Ed. 3-e, revision. - M.: Standard: NPC INFRA-M, 2014. - 624 p.

Naidenko V.V., Gubanov L.N. Environmental and economic monitoring of the environment: a training manual. - Nizhny Novgorod, 2003. -186 p.

Meshalkin V.P. Fundamentals of information system development and mathematical modeling of ecological systems : tutorial / V.P. Meshalkin, O.B. Butusov, A.G. Gnauk. - Moscow : INFRA-M, 2020. - 357 p.

Schreider Y.A. Intelligent systems and informatics // In the book: Intellect, man and computer. - Novosibirsk, 1994. - P. 72-90. Tsvetkov V.Y. Geoinformation systems and technologies. - Moscow: Finance and Statistics, 1998. - 288 p.

ST of RoK 34.005-2002. Information technology. Basic terms and definitions. - Introduced on 2004.01.01. - TC #34 on standardization "Information technologies". - Astana, 2002. - 36 p.

Shitikov V.K., Rozenberg G.S., Zinchenko T.D. Quantitative hydroecology: methods of system identification. - Tolyatti: Institute of Wolga Water Basin Ecology of the RAS, 2003. - 463 p.

Standard of the RoK 34.019-2002. Information technology. Processes of software life cycle. - Introduced on 2004.01.01. - TC № 34 on standardization "Information technology". - Astana, 2002. - 33 p.

<https://cleanbin.ru/problems/water-pollution>.

<https://vyvoz.org/blog/zagrjaznenie-vody/>.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 106-119
<https://doi.org/10.32014/2022.2518-1726.160>

УДК 004.021

**А.А. Муханова^{1*}, У.Т. Махажанова¹, Н.Д. Мархабатов¹, Б. Тасуов²,
Ж.Б. Ламашева¹**

¹Л.Н. Гумилева атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

²М.Х. Дулати атындағы Тараз өңірлік университеті, Тараз, Қазақстан.

E-mail: ayagoz198302@mail.ru

ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ТАЛДАУДА БҮЛДЫР ЛОГИКАНЫ ҚОЛДАНУ

Аннотация. Динамикалық өзгертін сыртқы ортада және қоғам жаһандануының нәтижесінде жаңа қасиеттері бар міндеттер пайда болады. Оларды шешу үшін тікелей тасымалдаушылары сарапшылар болып табылатын «сыртқы» білім қажет. Деректерді талдау тұрғысынан динамикалық жүйелер нақты объектілер мен процестердің әрекетін теориялық зерттеу үшін қызығушылық тудырады. Экономикалық ақпарат түрі мен өзгеру жылдамдығы бойынша ең алуан түрлі болып табылады, сондықтан берілген салада таңдау жасауға көмектесетін әдіс неғұрлым тұрақты аймақтарға қолданылуы мүмкін. Бұл жұмыс экономикалық ортада көп критериялды таңдау мәселелерін шешуге бағытталған. Бұлдыр жиындар теориясының математикалық аппаратын пайдалану негізінде шағын кәсіпорындардың несиелік қабілетін бағалау әдісі ұсынылған. Шешімдердің негізінде қабылданатын ережелер параметрлері бар логикалық формулалар түрінде қалыптасады. Кейбір параметрлер шағын кәсіпорынды талдау нәтижесінде алынады. Басқа параметрлер бұлдыр жиындар теориясы негізінде болжанады. Шағын кәсіпорынның несие қабілеттілігінің индексі деп аталатын ең жалпы түрде бір параметр болжанады, ол 0-ден 1-ге дейін өзгереді және бұның табиғи түсіндірмесі бар. Ұсынылған ғылыми тәсіл шағын бизнесті кәсіпорынды несиелендіру бойынша сараптамалық шешімдерді қолдау жүйелерін құру үшін негіз ретінде пайдаланылуы мүмкін.

Түйін сөздер: несиеге қабілеттілік, бұлдыр логика, лингвистикалық айналымы, логикалық ережелер, шешім қабылдау.

Қаржыландыру: Жұмысты ҚР ЖБФМ Ғылым комитеті, №AP0885497 гранты қолдады.

А.А. Муханова^{1*}, У.Т. Махажанова¹, Н.Д. Мархабатов¹, Б. Тасуов²,
Ж.Б. Ламашева¹

¹Евразийский национальный университет имени Л.Н. Гумилева,
Астана, Казахстан;

²Таразский региональный университет имени М.Х. Дулати, Тараз, Казахстан.
E-mail: *ayagoz198302@mail.ru*

ПРИМЕНЕНИЕ НЕЧЕТКОЙ ЛОГИКИ ПРИ АНАЛИЗЕ ЭКОНОМИЧЕСКИХ СИСТЕМ

Аннотация. В условиях динамично меняющейся внешней среды и в результате глобализации общества появляются задачи с новыми свойствами. Для их решения требуются «внешние» знания, непосредственными носителями которых являются эксперты. С точки зрения анализа данных динамические системы представляют интерес для теоретического исследования поведения реальных объектов и процессов. Экономическая информация, пожалуй, является самой разнообразной по типу и темпам изменения, поэтому, метод помогающий произвести выбор в данной отрасли, можно применять и по отношению к более устойчивым областям. Данная работа направлена на решение именно задач многокритериального выбора в экономической среде. Предложен метод оценки кредитоспособности малый предприятий, основанный на применении математического аппарата теории нечетких множеств. Правила, на основе которых принимаются решения формируются в виде логических формул, содержащих параметры. Одни параметры получаются в результате анализа деятельности предприятия. Другие параметры предсказываются на основе теории нечетких множеств. В самом общем виде предсказывается один параметр, называемый индекс кредитоспособности малый предприятий, меняющийся от 0 до 1 и имеющий естественную интерпретацию. Предлагаемый научный подход может быть использован в качестве основы для создания систем поддержки принятия экспертных решений при кредитовании предприятий малого бизнеса.

Ключевые слова: кредитоспособность, нечеткая логика, лингвистическая переменная, логические правила, принятие решений.

Финансирование: Работа поддержана Комитетом науки МВОН РК, грант № AP08855497.

**A. Mukhanova^{1*}, U. Makhazhanova¹, N. Markhabatov¹, B. Tassuov²,
Zh. Lamasheva¹**

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

²Taraz Regional University named after M.Kh. Dulaty, Taraz, Kazakhstan.

E-mail: *ayagoz198302@mail.ru*

APPLICATION OF FUZZY LOGIC IN THE ANALYSIS OF ECONOMIC SYSTEMS

Abstract. In a dynamically changing external environment and as a result of the globalization of society, tasks with new properties appear. To solve them, “external” knowledge is required, the direct carriers of which are experts. From the point of view of data analysis, dynamic systems are of interest for the theoretical study of the behavior of real objects and processes. Economic information is perhaps the most diverse in terms of type and rate of change, so a method that helps make choices in a given industry can be applied to more stable areas. This work is aimed at solving the problems of multi-criteria choice in the economic environment. A method for assessing the creditworthiness of small enterprises based on the use of the mathematical apparatus of the theory of fuzzy sets is proposed. The rules on the basis of which decisions are made are formed in the form of logical formulas containing parameters. Some parameters are obtained as a result of the analysis of the enterprise. Other parameters are predicted based on fuzzy set theory. In the most general form, one parameter is predicted called the index of creditworthiness of small enterprises, which varies from 0 to 1 and has a natural interpretation. The proposed scientific approach can be used as a basis for creating expert decision support systems for lending to small businesses.

Key words: classification creditworthiness, fuzzy logic, linguistic variable, logical rules, decision making

Funding: The work was supported by the Scientific Committee of the MHES of the RK, grant No. AP08855497.

Қысқашы. Соңғы уақытта шешім қабылдау әдістерін құруда міндеттегі сипаттамаларды адам тілі мен түсінігіне жуықтау және қалыптастыру бағытының дамуы байқалады. Объектіні бірнеше критерийлер бойынша сөзсіз бағалау күрделі, сарапшыға орташа бағаны іздеуге әкеледі.

Бастапқыда жаңа алгоритмдер қолданылатын негіздің өзін жіктеу қажет. Біздің жағдайда бұл - адамға немесе адамдар тобына экономикалық немесе басқа жүйенің әрі қарай қызметі тәуелді болатын шешім қабылдау барысындағы белгісіздіктер. Бұл жағдай, шешімдер қабылдау теориясында анықталған. Шешім қабылдау процесі кейбір жиындардан ең жақсы бір немесе бірнеше баламалар таңдау болып түсіндіріледі. Осындай шешім жасау үшін, мақсат және нұсқалардың баламалар жиынына бағалаулар жүргізілетін

критерийлерді айқын анықтау қажет. Осындай есеп шешімі әдісін таңдау қол жетімді ақпараттың сапасы мен санына байланысты (Чернов, 2006).

Кәсіпорындардың қаржы экономикалық жағдайын сипаттайтын ақпараттар, типі және өзгеру жылдамдығы бойынша әр түрлі болады, сондықтан, берілген салада таңдау жасауға көмектесетін әдісті басқа да тұрақты бағыттарда қолдануға болады, техникалық, медициналық. Осыған сүйене отырып, қарастырылған жұмыс экономикалық ортада көпкритерийлі таңдау есебін шешуге бағытталған. Кәсіпорындардың сыртқы ортасының маңызды айырмашылығы нарықтағы белгісіздіктер, оған сыртқы ортаның бақыланбайтын факторлары әсер етеді. Нарық жағдайында, сыртқы орта қолайлығы төмендеп, бәсекелестік күшейе түсті. Міндеттерде болатын белгісіздік – пікірлердің және сарапшы бағаларының айқын еместігімен, негізгі параметрлер мен есепті талдау жағдайында толық емес және айқын емес ақпараттармен сипатталады. Осылайша, ол шағын және орта кәсіпкерліктің несие қабілеттілігін бағалау есебінің күрделілігін айтарлықтай өсуіне әкеледі және көптеген факторлардың әсерінен пайда болады. Осы факторлардың тәжірибеде бірігуі әртүрлі белгісіздік түрлерінің кең ауқымын жасайды. Сондықтан көрсеткіштердің анық емес мәнін пайдалануға мүмкіндік беретін әдістерді қолдану қажеттілігі туындайды.

Міндеттерді шешу тиімділігі көбіне әдісті дұрыс таңдауға тәуелді. Шешім қабылдау әдістерінің әртүрлі критерийлерді қолдануға негізделген көптеген классификациялары бар. Шешім қабылдаудың көптеген белгілі әдістер мен тәсілдерінен ең жоғары қызығушылық беретіндер, көпкритерийлермен белгісіздіктерді есепке алатын, сонымен қатар түрлі өлшеу шкала типтері бар критерийлерлі әртүрлі типтегі баламалар жиынынан шешім таңдау. Осындай әдістерді экономикалық есептерді шешуде қолдануға болады:

- декомпозициялық әдістер пайдалылық күтілетін теориялар;
- иерархиялық талдау әдісі;
- бұлдыр жиындар теориясы.

Соңғы уақытта көпкритерийлі таңдау есебін шешім қабылдаушы тұлға (ШҚТ) қалауымен бұлдыр қатынастар әдістерімен шешу кең таралған. Ұқсас есептер, бір жағынан, қарапайым қатынастар есебіне қарағанда ресми жағынан жалпыланған, ал басқа жағынан олар іс жүзінде неғұрлым нақты, қарапайым, өмірде адамдардан басқа нұсқамен салыстырғанда бір нұсқаны сөзсіз қалауын талап ету тым қатаң. Көп жағдайда ШҚТ қандай да бір нұсқаға басқалармен салыстырғанда дәлелдері бар, сонымен қатар белгілі бір қарсылықтары бар. Бұл жағдайда ШҚТ кейбір дәрежеде сенімділік бар, яғни екі нұсқаның бірі қаланымы екеніне. Бұлдыр жиындар теориясы осындай ақпараттармен жұмыс жасауға негізделген.

Бұлдыр жиындарды қолданатын баламаларды көпкритерийлі таңдау әдістері:

- артықшылықты қатынастар;
- бұлдыр шешім;

- аддитивті айналдыру;
- максиминді айналдыру.

Әртүрлі механизмдерді ескере отырып әртүрлі нәтижелер алу толығымен мүмкін жағдай. Көп жағдайда дәрежелендірудің жалпы реті бір жолмен жүргізіледі, функцияның тиістілік мәні бір-бірінен өзгеше болады.

Жоғарыда айтылғандарды қорытындылай келе, шағын және орта бизнесті кәсіпорындардың несиелендірілуінің қабілеттілігін бағалау мәселесі өзекті, тәжірибелік маңызды, сонымен қатар жеткіліксіз әзірленген. Жүйе жұмыс жасайтын орта тұрақты емес, себебі жүйелердің жұмыс істеуінің сыртқы және ішкі жағдайларының өзгеруі несиелендіру шарттары мен несиелендірілуінің қабілеттілігін сипаттайтын параметрлерді қайта қарау қажеттілігіне алып келеді. Осы себепті шағын және орта бизнесті кәсіпорындарды несиелендіру орындылығын талдауды, жоғары дәрежедегі белгісіздік жағдайында шешілетін есеп ретінде қарастыру қажет. Осылайша, белгісіздіктің бар барлық түрін есепке алатын теорияны шешім қабылдау үшін қолдану қажеттілігі туындайды.

Әдебиеттерге шолу. Көп критерийлі оңтайландыру, динамикалық бағдарламалау және бұлдыр логика сұрақтары мамандардың үлкен тобымен зерттелді. Осы мәселемен айналысқан ғалымдар арасында, бұлдыр логика саласындағы маңызды әзірлемелер – Л.А. Заде көпкритерийлі оңтайландыру мәселесі – Р. Штойер, динамикалық бағдарламалаудың негізгі жұмыстары – Р.И. Беллман жұмыстарын ерекшелеуге болады.

Соңғы уақытта анық емес және біртегіс емес жүйелерді зерттеу үшін тиімді аспап ретінде бұлдыр жиындар теориясы тәсілін қолданылуда. Бұлдыр жиындар теориясының дамуына, теория және практикада бұлдыр модельдер құруға үлкен үлес қосқан шетел ғалымдары: Л. Заде, Е. Мамдани, М. Сугэно, Д. Дюба, Т. Тэрано, А. Кофман, С.М. Ковалев, О.В. Недосекин, М.В. Семухин және т.б. Қазіргі уақытта экономикалық зерттеулерде анық емес логикалық құралдар кеңінен қолданылады. Банктер мен фирмалардың банкроттықты болжау мәселесін шешу үшін статистикалық және болжамдық әдістерді қолдану бойынша 1968–2005 жылдары атқарылған жұмыстарға жан-жақты шолу жасалған (Кутар және б., 2007). Жаңа сауда стратегиясын және инвесторларға портфолиосын басқаруға көмектесу үшін модификацияланған жиі қолданылатын және сирек қолданылатын техникалық көрсеткіштер арасындағы «қоспасын» қолданатын анық емес қысқа мерзімді сауда жүйесін ұсынылған (Choumouziadi және б., 2016). Әлеуметтік-экономикалық жүйелер модельдері үшін анық емес когнитивтік карталардың қолдану мүмкіндігі зерттелді (Carvalho және б., 2009). Кипрдегі мұнай және газ өндірісінің әлеуметтік-экономикалық әсерін бағалаудың когнитивтік моделі жасалынды (Neocleous және т.б., 2011). Үздіксіз модельдер теориясы тұрғысынан Л.А. Заде өте ерекше жағдай. Дегенмен, бұл қызықты, себебі ол техникалық, табиғи және гуманитарлық салаларда көптеген қолданбаларды тапты [6] (Zadeh және б., 1996).

Л.А. Заде логикасы – қарапайым және табиғи классикалық емес логикалардың бірі. Жоғарыда айтылғандай, логика формулаларының мәндері Л.А.

Заде сегментте жатыр. «Үздіксіз модельдер теориясы» әдебиетінде жалпы жағдайды, атап айтқанда, формулалардың мәндері топологиялық кеңістікте жатқанда қарастырды (Keisler және т.б., 1966). Зерттелетін негізгі жағдай топологиялық кеңістік ықшам және Хаусдорф, ал логикалық қосылғыштар мен кванторлар үздіксіз функциялар болғанда. Бір қызығы, классикалық логиканың көптеген конструкциялары мен теоремалары (дәлірек айтқанда, модель теориясы) осы жағдайды жалпылайды.

Материалдар мен әдістер. Шағын және орта бизнесті кәсіпкерлікті – қарыз алушыны несиелендіру мақсаттылығын талдау есебін баламаларды таңдаудың көпкритерийлі есеп шешу математикалық әдістерімен шешуге болады.

Бағаланатын сипаттамалардың (критерийлер жиыны) анықталған жиыны және кейбір баламалар жиыны (баламалар ретінде несиелендіру шарттары) бар болғандықтан бұл есепті баламаларды таңдаудың көпкритерийлі есебі деп қарастыруға болады.

Бұл бөлімде баламаларды көпкритерийлі таңдау есебін шешудің қолданыстағы әдістері қарастырылады, қарастырылған әдістердің артықшылықтары мен шектеулері көрсетіледі. Сонымен қатар, шағын және орта бизнесті кәсіпкерлікті несиелендіру мақсаттылығын талдауда осындай әдістерді қолдану негізделеді.

Қолданыстағы әдістердің ішінде шартты логикалық шешім ережелері (өндірістік ереже) өте кең таралған.

Есеп қойылымын тұжырымдау.

U – элементтер жиыны, тиістілік дәрежесі $[0,1]$ аралығындағы сандар болатын A – оның бұлдыр ішкі жиыны. A_j – ішкі жиынының мәні лингвистикалық айнымалы X болады.

Шешімдер жиыны – x_1, x_2, \dots, x_n критерийлер жиынтықтарымен сипатталады делік, яғни u_1, \dots, u_n негізгі жиынына сәйкес берілген лингвистикалық айнымалылар. Мысалы, x_1 айнымалысы «аймақтың даму динамикасы» мәні «Жоғары», ал x_2 «нарықтың осындай тауарларға сұранысы» мәні «Орташа». Сәйкес мәнді бірнеше критерийлер жиыны қанағаттанарлық балама туралы шешім қабылдаушы тұлға бейнесін сипаттайды (Андрейчиков және б., 2000).

Бірінші кезеңде барлық көрсеткіштер бағаланады және барлық мүмкін болатын нәтижелер анықталады. Осыған сәйкес, төменде көрсетілген толық шартты ережелер жиынтығы қалыптасады (Соколова және б., 2013).

Ережеі :«Егер 1» және «Егер 2» және ... «Егер M», онда «Шешім X». (3.6.1.1).

R 1: Егер Көрсеткіш 1 = «Төмен» және Көрсеткіш 2 = «Төмен» және ... Көрсеткіш N = «Төмен», Онда «Шешім 1».

R 2: Егер Көрсеткіш 1 = «Орташа» және Көрсеткіш 2 = «Төмен» және... Көрсеткіш M = «Орташадан төмен», Онда «Шешім 2».

R n: Егер Көрсеткіш 1 = «Жоғары» және Көрсеткіш 2 = «Жоғары» және ... Көрсеткіш M = «Жоғары», Онда «Шешім N».

Келесі қадам сол жақ ережелердің μ_M тиістілік дәрежесін есептеу.

μ_M i -ші шешімінің тиістілік дәрежесі төмендегі формула бойынша анықталады:

$$\mu_{Mi} = \min(\mu_{x1}(u), \mu_{x2}(u), \dots, \mu_{xn}(u))$$

Несие қабілеттілігін бағалаудың Шешімі мүмкін нұсқалары өздерінің тиістілік функциялары B_i болатын бұлдыр жиындар түрінде ұсынылады, мысалы: төмен, орташадан төмен, орташа, орташадан жоғары, жоғары.

Осылайша, критерийлер жинағын қолданып жоғарыда сипатталған ережелерді төмендегідей жазуға болады:

R 1: Егер $X = A1$, Онда $S = B1$,

...

R 2: Егер $X = A5$, Онда $S = B5$.

Баламаларды таңдау кезінде олардың әрқайсының қанағаттануы табылады және сәйкес келетін нүктелі баға есептеледі. Ең жақсы балама – ең жоғары мәнді балама болып есептеледі.

Жалпы алғанда бұлдыр логикалық шешім әдістемесі төрт кезеңнен тұрады: бұлдырлықты енгізу (фаззификация), бұлдыр шешім, композиция және айқындыққа келтіру (дефаззификация).

Қарастырылған бұлдыр шешім кезеңдері біртекті емес түрде жүзеге асырылуы мүмкін, осылайша, бұлдыр шешімнің әр кезеңінің жүзеге асуының нақты жолдарын сол немесе басқа бұлдыр шешім алгоритмін таңдау анықтайды. Қазіргі заманда нақты техникалық есепке тәуелді бұлдыр шешім алгоритм әдістері мен критерийлерін таңдау ашық сұрақ болып қалуда. Қазіргі уақытта бұлдыр шешім жүйелерінде төмендегі алгоритмдер жиі қолданылады:

1. Мамдани (Mamdani) алгоритмі.
2. Цукамото (Tsukamoto) алгоритмі.
3. Ларсен алгоритмі.
4. Сугено алгоритмі және т.б. (Леоненков және б., 2003).

Мысал ретінде ең көп таралған бұлдыр шешім алгоритм модификацияларын қарастырамыз, білім қорында тек екі бұлдыр ережелер түрі бар деп болжай отырып:

Ереже₁: егер x бар A_1 және y бар B_1 , онда z бар C_1 ,

Ереже₂: егер x бар A_2 және y бар B_2 , онда z бар C_2 .

мұнда x, y – енгізілетін айнымалылар атаулары; z – шығатын айнымалы атауы.

$A_1, A_2, B_1, B_2, C_1, C_2$ – тиістілік функцияларымен берілген айнымалыларды бағалайтын лингвистикалық термдар, сонымен қатар z_0 айқын мәнін келтірілген ақпарат және x_0, y_0 айқын мәндері негізінде анықтау қажет.

Ресми түрде Мамдани алгоритмі төмендегідей сипатталуы мүмкін:

1. Бұлдырлық. Әр ережеге алғышарттар үшін шындық дәрежесі анықталады – $A_1(x_0), A_2(x_0), B_1(y_0), B_2(y_0)$

2. Бұлдыр шешім. $\alpha_1 = A_1(x_0) \wedge B_1(y_0)$; $\alpha_2 = A_2(x_0) \wedge B_2(y_0)$ ережелерден әрқайсына алғышарт үшін «қиылысу» деңгейлері табылады, мұнда \wedge арқылы логикалық минимум операциясы белгіленген. «қиылған» тиістілік функциялары $C'_1 = (\alpha_1 \wedge C_1(z))$; $C'_2 = (\alpha_2 \wedge C'_2(z))$ табылады.

3. Композиция. Логикалық максимум (\max , \vee) операциясын қолданып табылған қиылған функциялар біріктіріледі және шығатын айнымалы тиістілік функциясы $\mu_\Sigma(z) = C(z) = C'_1(z) \vee C'_2(z)$; үшін қорытынды бұлдыр ішкі жиыны алынады.

4. Айқындыққа келтіру (z_0 табу үшін) ереже бойынша, центроидты әдісімен өткізіледі – шығатын айнымалы мәні айқын екендігі $\mu_\Sigma(z)$ қисығы үшін ауырлық орталығы ретінде анықталады (Mamdani және б., 1975).

Сугено алгоритмі

Сугено және Такаги ережелер жиынын төмендегі түрде қолданды (Takagi және б., 1985):

Ереже₁: егер x бар A_1 және y бар B_1 , онда $z_1 = a_1x + b_1y$,

Ереже₂: егер x бар A_2 және y бар B_2 , онда $z_2 = a_2x + b_2y$.

Алгоритмнің сипатталуы:

1. Бұлдырлыққа келтіру Мамдани алгоритміндегідей болады.

2. Бұлдыр шешім. $\alpha_1 = A_1(x_0) \wedge B_1(y_0)$; $\alpha_2 = A_2(x_0) \wedge B_2(y_0)$ табылады және жеке шешім ережелері табылады:

$$\dot{z}_1 = a_1x_0 + b_1y_0$$

$$\dot{z}_2 = a_2x_0 + b_2y_0$$

3. Нәтиже айнымалысының айқын мәні анықталады:

$$z_0 = \frac{a_1\dot{z}_1 + a_2\dot{z}_2}{a_1 + a_2}$$

Қарастырылған жағдайда Мамдани алгоритмін тікелей қолдану бірқатар қиындықтардың туындауына әкеледі. Қарастылған есепте қолданылған лингвистикалық айнымалылар, жалпы жағдайда әртүрлі әмбебап жиындарда анықталған болуы мүмкін, сондықтан дәстүрлі Мамдани алгоритмінде ереженің шартты бөлімінде критерийлерді айналдыру тиістілік функциясы орнына олардың синглтондары қолданылады. Егер техникалық жүйелерді басқару үшін бұл маңызды болмаса, қарастырылып отырған есепте мұндай ауыстыру дұрыс емес нәтижелерге әкеледі. Сонымен қатар, бұлдыр критерийлер қиылыстарын қалыптастыру үшін минимум операциясын қолдану, шешім мәні кейбір критерийлер бойынша ең төменгі бағамен анықталады, мұнда да берілген есеп бойынша дұрыс емес нәтиже болуы мүмкін.

Қорытындылай келе бұлдыр логикалық жүйелердің негізгі артықшылықтары:

- бұлдыр енгізу мәліметтерімен жұмыс істеу мүмкіндігі, мысалы, уақыт аралығында үздіксіз өзгеріп отыратын мәндер (динамикалық есептер), біртекті беру мүмкін емес мәндер;

- бағалау критерийлерін бұлдыр қалыптастыру және салыстыру: «көпшілігі», «мүмкін», «басым» және т.б. критерийлармен жұмыс істеу мүмкіндігі;

- сапалық бағаларды енгізу мәліметтері ретінде өткізу мүмкіндігі, сонымен қатар шығу нәтижелері: мәліметтер мәнімен ғана емес, олардың шындық дәрежесімен және оның таралуымен жұмыс істейді;

- күрделі динамикалық жүйелерді жылдам модельдеу мүмкіндігі және олардың берілген дәлдік дәрежесін салыстармалы талдау мүмкіндігі.

Бұлдыр жүйелердің кемшілігі:

- бұлдыр жүйе құрудың стандартты әдістердің болмауы;

- бұлдыр жүйенің қолданыстағы әдістерімен математикалық талдау жасау мүмкіндігінің болмауы;

- ықтималдықпен салыстырғанда бұлдыр логиканы қолдану есептеудің дәлдігін арттырмайды.

Кәсіпорындардың несиеге қабілеттілігін бағалау көрсеткіштерін талдау

Қарыз алушы кәсіпорындардың несиеге қабілеттілігін бағалау процесінде сапалық параметрлерге ерекше мән берілуі қажет. Солардың бірі кәсіпорын жататын сала болып табылады. Саланың даму бағытын талдау сыртқы факторлардың өзгеру қатерін болжауға мүмкіндік береді, осы факторлар кәсіпорынның табыстылығына, активтерінің сапасына және ең бастысы қарыз міндеттемелерін атқару мүмкіндігіне елеулі әсер етеді. Аталған қатерлер экономиканың барлық салаларында кездеседі және олардан бірде-бір кәсіпорын айналып өте алмайды.

Кәсіпорындардың несиеге қабілеттілігін бағалау көрсеткіштері сапалық және сандық болып бөлінеді. Сандық көрсеткіштерге қаржылық коэффициенттерді, ал сапалық көрсеткіштерге кәсіпорынның сапа мен бағыт ішіндегі бәсекеге қабілеттілік деңгейі жатады.

Қарыз алушының несиеге қабілеттілігін бағалауда ескерілетін белгілі бір саланың спецификасының сапалық көрсеткіштері:

Саланың даму динамикасы – салалақ статистикалық мәліметтер мен аймақтық департаменттер мәліметтері негізінде саланың даму динамикасы сарапталады.

Саланың даму перспективалары–саланың даму сценарилері (оптимисті және пессимисті) қолданылады.

Сала нарығының осындай өнімге (қызметке) сұранысы –1. Өнімге ұсыныспен сұраныс, өнім ассортименті мыналардың көмегімен сарапталады: а) ассортимент бойынша жоспарды орындауды бағалау; б) сала бойынша бағаларды саралау; в) өнімнің орташа бағасына оның сапасының әсер ету көрсеткішін саралау. 2. Бәсекеге қабілеттілікті сараптау салмақ коэффициенттері негізінде интегралды көрсеткіштер көмегімен жүзеге асырылады (Заболоцкая және б., 2009).

Қаржы коэффициенттерін таңдауда төмендегі критерийлерді қолдану қажет:

- коэффициенттер саны өте үлкен болмау керек;

- коэффициенттер бір-бірін қайталамау қажет;

- коэффициенттер қарыз алушының қаржылық жағдайын неғұрлым жоғары көрсету қажет;

Шағын кәсіпкерлікті несиелеудің орындылығын анықтау үшін есептелген көрсеткіштер топтарының жиынтығы: салалық және аймақтық ерекшеліктер, шағын кәсіпкерлік қызмет және қызмет көрсету және сауда салаларына тән қаржылық-экономикалық көрсеткіштер қарастырылды.

Нәтижелер. Бұл бөлімде ұқсас көрсеткіштер $[0,1]$ интервалы аралығында көрсетілетіні, көрсеткіштерді бірыңғайландыру және Әрбір параметрмен бір «осы көрсеткіштің қолайлылық дәрежесі» лингвистикалық айнымалысын байланыстыру жағдайы Makhazhanova жұмысында сипатталған.

Бастапқы көрсеткіштер $[0,1]$ аралығында көрсетіледі. Параметрлер жиыны лингвистикалық айнымалылар жиыны түрінде ұсынылады. Көп қажетті ақпараттар әдебиеттерде бар (Махажанова және б., 2020).

Жалпы жағдайда, кез келген параметр кейбір интервалда нақты сандар мәнін қабылдайды деп есептейміз. Қарапайым жағдайда, әрбір параметрге «осы көрсеткіштің қолайлылық дәрежесі» деп атауға болатын, бір лингвистикалық айнымалы сәйкестендіріледі. Нақты жағдайда, параметрге $[0,1]$ интервалы аралығын қабылдай алатындай нақты сан беру қажет. Осы процесс бірыңғайландыру процесі деп аталады.

Төменде келтірілген функциялармен бірге «қолайлылық дәрежесі» композициясын алып, сәйкесінше жаңа лингвистикалық айнымалылар мәнін енгіземіз: «параметр мәні төмен» (L); «параметр мәні орташа» (M); «параметр мәні жоғары» (H). Осындай түрдегі лингвистикалық айнымалылар техникалық жүйелерде жиі қолданылады. Оданда көбірек санды градациялар қолданылады: орташадан төмен, орташадан жоғары, нөлге жақын және т.б.

Әрбір лингвистикалық айнымалымен өзінің тиістілік функциясы байланыстырылады. Сайып келгенде, тиістілік функциялар түрін таңдау, баламалар сияқты, бағаланатын критерийлер үшін де сараптамалық артықшылықтар негізінде анықталады. Айта кету керек, көрсеткіштер (параметрлер) ерікті мән алады. Дәлірек айтқанда, әрбір көрсеткіш белгілі бір аралықта өзгереді. Бұдан әрі көрсеткіштерді біріктіруге болады, яғни, $[0,1]$ интервалы арасында көрсету (Тусупов және б., 2020).

Бағалауда сарапшы сенімсіздігі бағаланатын параметр мәнінің оңтайлы мәнен ауытқуының артуымен өседі. Сонымен қатар, көп жағдайда берілген сенімсіздік сызықты емес өседі. Бірақ сызықты емес тиістілік функцияларын қолдану математикалық есептеулерді және график құруды айтарлықтай қиындатады.

Осы себепті, бұл жұмыста бастапқы тиістілік функциялары ретінде үшбұрышты және трапециялық тиістілік функциялары қолданылады, бұл ең алдымен кейінгі есептеулер орындау мен график құру қарапайымдылығымен байланысты. Тиістілік функциялары үзік сызықты деп айтуға болады.

Таңдалған бағалау критерийлер жиынтығы сол немесе басқа баламаға сәйкестік дәрежесін анықтау – кейінгі несиелендіру үшін оңтайлы таңдалған

шарттар жиынтығы бар, ең қолайлы несиелендіру схемасын таңдау кезіндегі кілттік фактор болады.

Жоғарыда айтылғандай, параметрлер (көрсеткіштер) ерікті мәндер қабылдайды. Дәлірек айтқанда, әрбір параметр өзіне тән анықталған интервалда өзгереді. Ары қарай, параметрлерге бірыңғайландыру жүргізілді, яғни сәйкес интервалдар $[0,1]$ аралығында көрсетілді деп есептейміз.

Шешім қабылданатын ережелердің логикалық формула кескінін құру

Қарыз алушыны несиелендіру немесе несиелендіруден бас тарту шешімін қабылдау үшін A жиынының a_j критерийі бойынша X жиынының x_i көрсеткішінің кескінін құру қажет.

Нәтижесінде критерийлер жиынын есепке ала отырып әрбір x_i көрсеткіші үшін өз несие қабілеттілік X_i^* мәнін табу мүмкіндігі пайда болады.

Қабілеттілік мәнін табу рәсімі X жиынының барлық көрсеткіштері үшін орындалады, және алынған мәндер қосылады, қорытынды шешім F мәнін қалыптастырып. F мәні несиелендіру немесе бас тарту шешімін қабылдауға мүмкіндік береді, сонымен қатар қорытынды F негізінде оңтайлы несиелендіру схемасы таңдалады .

Шешім қабылданатын ережелер параметрлері бар логикалық формулалар түрінде қалыптастыралды:

$$F_j = \sum_{i=1}^n \omega_i X_i^*$$

мұнда F_j – j -ші шағын кәсіпорын несие қабілеттілігі;

X_i^* – шағын кәсіпорынды талдау нәтижесінде алынған көрсеткіштер;

ω_i – көрсеткіштердің салмақ дәрежесі Саати формуласымен есептеледі.

Бұлдыр жиындар теориясына сәйкес F бұлдыр жиындар қуатына сәйкес келеді, F неғұрлым жоғары болса, шағын кәсіпкерліктің несие қабілеттілігі сондай жоғары.

Ұсынылған баламаларға сәйкес (L, LM, M, NM, N) жоғары несие қабілеттілік қуаты: төмен, орташадан төмен, орташа, орташадан жоғары, жоғары.

Әрбір қорытынды шешімге сәйкес өз несиелендіру шарттары сәйкес болады, яғни қорытынды шешімді есептегеннен соң оңтайлы несиелендіру схемасы анықталады.

Талқылау. Шағын кәсіпкерлікті несиелеудің орындылығын анықтау үшін есептелген көрсеткіштер топтарының жиынтығы қарастырылды: салалық және аймақтық ерекшеліктер, шағын кәсіпкерлік қызмет және қызмет көрсету және сауда салаларына тән қаржылық-экономикалық.

Қаржылық-экономикалық көрсеткіштердің жиынтығы несиелік сарапшының қалауына байланысты немесе қарыз алушы ұсынған кіріс ақпаратына сәйкес өзгертілуі мүмкін, олар қаржылық есептілік немесе оңайлатылған салық салу жүйесінің деректері ретінде пайдаланылуы мүмкін.

x_i көрсеткішінің деңгейін жіктеу кезінде келесі мәндер пайдаланылады: $a_1 = L$ – L (төмен) – төмен деңгей; $a_2 = LM$ (төмен орта) - орташадан төмен;

$a_3 = M$ (орташа) - орташа; $a_4 = HM$ (жоғары орта) - орташадан жоғары; $a_5 = H$ (жоғары) - жоғары.

Сайып келгенде, мүшелік функциялардың түрін таңдау, баламалардың өзі де, бағаланатын критерийлер де сарапшылардың қалауы негізінде анықталады.

Сарапшының бағалаудағы белгісіздігі бағаланатын параметр мәнінің оңтайлы мәннен ауытқуының жоғарылауымен өсетіні анық. Оның үстіне, көп жағдайда бұл белгісіздік сызықты түрде өспейді. Дегенмен, сызықты емес мүшелік функцияларды пайдалану математикалық есептеулер мен графикалық конструкцияларды айтарлықтай қиындатады.

Осы себептерге байланысты бұл жұмыста үшбұрышты және трапеция тәрізді мүшелік функциялары бастапқы мүшелік функциялар ретінде пайдаланылады, бұл ең алдымен кейінгі есептеулер мен графикалық конструкцияларды орындаудың қарапайымдылығына байланысты. Мүшелік функцияларды бөліктік сызықтық деп айта аламыз. Таңдалған бағаланатын критерийлер жиынтығының сол немесе басқа баламаға сәйкестік дәрежесін анықтау кейіннен несиелеу үшін оңтайлы таңдалған шарттар жиынтығы бар неғұрлым қолайлы несиелеу схемасын таңдаудың негізгі факторы болып табылады.

Сонымен қатар, атап айту керек, қарыз алушының несие қабілеттілігін бағалау есебінде бұлдыр жиындар теориясы математикалық аппаратын қолдану динамикалық сапада, бүкіл несиелендіру мерзімі ішінде қарыз алушы күйін талдау процессін қажетті жиілікте қайталауға мүмкіндік береді. Бұл тек қана несиелендіру мерзімі ішінде қарыз алушы күйіне бақылау жүргізу емес, сонымен қатар қолданыстағы несиелік қауіптілікті мерзімді бағалау жүргізуді қарастыратын заңнама талаптарына сәйкестігін қамтамасыз етеді.

Қорытынды. Қарыз алушы – шағын және орта кәсіпорындардың несие қабілеттілігін бағалау – мұқият жүргізілуі керек, шағын кәсіпорынның даму деңгейін, қызмет ерекшеліктерін, сонымен қатар белгілі бір қаржылық нәтиже алған кезде белгісіздік факторларын ескеру қажет. Шартты бұлдыр ережелерге негізделген бұлдыр жиындар теориясы аппаратын қолдану жоғарыда келтірілген шектеулерді, белгісіздіктерді ескеруге мүмкіндік береді, бұл қазіргі кезде қолданылатын әдістермен салыстырғанды олардың сөзсіз артықшылығы болып табылады, қарыз алушыны жан-жақты дұрыс бағалауды қамтамасыз етеді.

Қолданыстағы әдістермен салыстырғанда ұсынылған тәсіл шағын бизнестің несиелік қабілетін талдау кезінде қабылданған шешімдердің негізділігінің жоғарылауымен сипатталады, бұл бұлдыр жиындар теориясының математикалық аппаратын қолданумен байланысты. Қолданыстағы әдістерден ерекшелігі бұлдыр жиындар теориясына негізделген барлық әдістер, ортақ қасиеттерге ие:

- бұлдыр модельді шешім қабылдау әдістері жеке критерийлер бойынша баламаларға ыңғайлы және сапалы бағалау жүргізеді. Басқа әдістермен салыстырғанда жаңа баламаларды қосу алдыңғы дәрежеленген жиындар

ретін өзгертпейді. Критерийлар бойынша баламаларды бағалау кезінде лингвистикалық бағалау, сонымен қатар критерийлердің тиістілік функциясын қолданып нүктелі бағалау негізінде бағалау мүмкіндігі.

- бұлдыр модельді қолданып көпкритерийлі тандаудың негізгі мәселесі критерийлер арасындағы қатынас туралы ақпаратты ұсыну және интегральды бағаларды есептеу жолы болып табылады. Өртүрлі тәсілге негізделген әдістер түрлі нәтижелер береді. Әрбір тәсілдердің өзінің шектеулері мен ерекшеліктері бар. Ақпаратты ұсыну үшін ең ауқымды мүмкіндіктерді эвристикалық тәсілдер береді.

Information about the authors:

A. Mukhanova – PhD, Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: ayagoz198302@mail.ru, <https://orcid.org/0000-0003-3987-0938>;

U. Makhazhanova – PhD, Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: makhazhan.ut@gmail.com, <https://orcid.org/0000-0002-5528-8000>;

N. Markhabatov – PhD, Department of Algebra and Geometry, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: nur_24.08.93@mail.ru, <https://orcid.org/0000-0002-5088-0208>;

B. Tassuov – Candidate of technical sciences, Department of Physics and Information Technology, Taraz Regional University named after M.Kh. Dulaty, Taraz, Kazakhstan, E-mail: bolat_ktn@mail.ru, <https://orcid.org/0000-0002-2000-6720>;

Zh. Lamasheva – PhD, Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: zhanarlb@gmail.com, <https://orcid.org/0000-0002-9535-2636>.

ӘДЕБИЕТТЕР:

Keisler H.J., Chang C.C. Continuous Model Theory, Princeton University, 1966.

Takagi T., Sugeno M. Fuzzy Identification of Systems and Its Applications to Modeling and Control. IEEE Trans. Systems, Man, and Cybernetics. 1985, 15, №1, 116-132.

Тусупов Д.А., Муханова А.А., Махажанова У.Т. Шағын бизнесті кәсіпорынның несиеге қабілетін бағалаудың математикалық моделі. Вестник КазНУ. 2020, №2(138), 728-731.

Chourmouziadis K., Chatzoglou P.D. An intelligent short term stock trading fuzzy system for assisting investors in portfolio management. Expert Systems with Applications, 2016, 43, 298-311. <https://doi.org/10.1016/j.eswa.2015.07.063>

Carvalho J.P., Tome J.A.B. Rule based fuzzy cognitive maps in socio-economic systems. 2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference, IFSA-EUSFLAT 2009 – Pro-ceedings, 2009, 1821-1826.

Mamdani E.H. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*. 1975, 7, 1-13.

Makhazhanova U.T., Murzin F.A., Mukhanova A.A., Abramov E.P. Fuzzy logic of Zadeh and decision-making in the field of loan. *Journal of theoretical and applied Information Technology*, 2020, 98 (06), 1076-1086.

Neocleous C., Schizas C., Papaioannou M. Fuzzy cognitive maps in estimating the repercussions of oil/gas exploration on politico-economic issues in Cyprus. *IEEE International Conference on Fuzzy Systems*, 2011, 6007655, 1119-1126.

Заболоцкая В.В., Аристархов А.А. Повышение эффективности оценки кредитоспособности предприятий малого бизнеса в коммерческом банке. *Банковское дело*, 2009, №4, 99-111.

Zadeh L.A. Shadows of fuzzy sets. *Advances in Fuzzy Systems – Applications and Theory Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems*, 1996, 51-59. https://doi.org/10.1142/9789814261302_0004

Ravi Kumar, P., Ravi, V. Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *Eu-ropean Journal of Operational Research*, 2007, 180 (1), 1-28 <https://doi.org/10.1016/j.ejor.2006.08.043>.

REFERENCES:

Keisler, H.J., Chang, C.C. *Continuous Model Theory*, Princeton University, 1966.

Takagi T., Sugeno M. Fuzzy Identification of Systems and Its Applications to Modeling and Control. *IEEE Trans. Systems, Man, and Cybernetics*. 1985, 15, №1, 116-132.

Tusupov D.A., Mukhanova A.A., Makhazhanova U.T. Shagyn biznesi kasiporynnyn nesiege kabiletin bagalaudyn matematik modeli. *Bulletin of Kazntu*. 2020, №2(138), 728-731.

Chourmouziadis, K., Chatzoglou, P.D. An intelligent short term stock trading fuzzy system for assisting investors in portfolio management. *Expert Systems with Applications*, 2016, 43, 298-311. <https://doi.org/10.1016/j.eswa.2015.07.063>

Carvalho, J.P., Tome, J.A.B. Rule based fuzzy cognitive maps in socio-economic systems. 2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference, IFSA-EUSFLAT 2009 – Pro-ceedings, 2009, 1821-1826.

Mamdani E.H. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*. 1975, 7, 1-13.

Makhazhanova, U.T., Murzin, F.A., Mukhanova, A.A., Abramov E.P. Fuzzy logic of Zadeh and decision-making in the field of loan. *Journal of theoretical and applied Information Technology*, 2020, 98 (06), 1076-1086

Neocleous, C., Schizas, C., Papaioannou, M. Fuzzy cognitive maps in estimating the repercussions of oil/gas exploration on politico-economic issues in Cyprus. *IEEE International Conference on Fuzzy Systems*, 2011, 6007655, 1119-1126.

Zabolotskaya V.V., Aristarkhov A.A. Improving the efficiency of assessing the creditworthiness of small businesses in a commercial bank. *Banking*, 2009, No. 4, 99-111.

Zadeh L.A. Shadows of fuzzy sets. *Advances in Fuzzy Systems – Applications and Theory Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems*, 1996, 51-59. https://doi.org/10.1142/9789814261302_0004

Ravi Kumar, P., Ravi, V. Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *Eu-ropean Journal of Operational Research*, 2007, 180 (1), 1-28 <https://doi.org/10.1016/j.ejor.2006.08.043>.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 120-133
<https://doi.org/10.32014/2022.2518-1726.161>

UDC 004.056.5

**N. Seilova*, A. Batyrgaliyev, Zh. Dzhangozin, D. Baibatchayeva,
N. Nurgabylov**

Kazakh National Research Technical University named after K.I. Satpayev,
Almaty, Kazakhstan.
E-mail: *seilova_na@mail.ru*

METHOD FOR ASSESSING THE QUALITY OF MASKING NOISE INTERFERENCES

Abstract. Currently, the number of various means of computer technology designed for processing, storing, and transmitting information is increasing. One of these threats is the presence of technical channels of information leakage arising from informative (dangerous) spurious electromagnetic radiation from technical means of its processing, storage and transmission. To protect information from leakage through the channels of spurious electromagnetic radiation and interference, systems of spatial electromagnetic noise are widely used.

However, the quality of the masking noise interference of these systems is not always assessed, both during their development, design, and production, and in confirming their compliance with information security requirements. As a rule, they are limited to checking the compliance of the operating frequency range, spectral density, type of radiated interference, etc. The noise quality entropy factor is used as a measure of the quality of the masking noise generated by the noise generator. A sufficient number of studies have been devoted to the creation of generators of spatial electromagnetic noise.

Thus, the study of the quality characteristics of generators of spatial electromagnetic noise is relevant. A technique for estimating the entropy noise quality factor using a spectrum analyzer and a digital storage oscilloscope is proposed. The procedure for calculating the entropy noise quality factor from the statistics of the instantaneous values of the noise signal amplitudes measured by a digital oscilloscope is given.

Key words: masking noise signal; noise generator; spectrum analyzer; entropy coefficient.

**Н.А. Сейлова*, А.Б. Батыргалиев, Ж.А. Джангозин, Д.А. Байбатчаева,
Н. Нұрғабылов**

Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті,
Алматы, Қазақстан.
E-mail: *seilova_na@mail.ru*

ШУ КЕДЕЛДЕРІН БҮРКЕУДІҢ САПАСЫН БАҒАЛАУ ӘДІСТЕМЕСІ

Аннотация. Қазіргі уақытта ақпаратты өңдеуге, сақтауға және беруге арналған компьютерлік техниканың әртүрлі құралдарының саны артып келеді. Осы қауіптердің бірі оны өңдеудің, сақтаудың және берудің техникалық құралдарынан ақпараттық (қауіпті) жалған электромагниттік сәулеленуден туындайтын ақпараттың ағып кетуінің техникалық арналарының болуы болып табылады. Бұл ақпаратты жалған электромагниттік сәулелену және кедергі арналары арқылы ағып кетуден қорғау үшін кеңістіктік электромагниттік шу жүйелері (шу генераторлары) кеңінен қолданылады.

Алайда, бұл жүйелердің шу кедергілерін бүркемелеу сапасы оларды әзірлеу, жобалау және өндіру барысында да, олардың ақпараттық қауіпсіздік талаптарына, күнделікті жұмысына сәйкестігін растау кезінде де бағалана бермейді. Әдетте, олар жұмыс жиілігі диапазонының, спектрлік тығыздықтың, сәулелену кедергілерінің түрін, жұмыс режимдері мен шарттарын, сондай-ақ санитарлық нормаларды сақтау қажеттілігін және т.б. сәйкестігін тексерумен шектеледі. Шу сапасының энтропия коэффициенті шу генераторы тудыратын маска шуының сапасының өлшемі ретінде пайдаланылады. Кеңістіктік электромагниттік шудың генераторларын құруға зерттеулердің жеткілікті саны арналды. Бірақ қазіргі уақытта шудың сапасын бағалау әдістері жоқ, олардың негізінде дайын техникалық шешімдер бар.

Осылайша, кеңістіктік электромагниттік шу генераторларының сапалық сипаттамаларын зерттеу өзекті болып табылады. Жұмыста спектр анализаторы мен сандық сақтау осциллографы арқылы шудың энтропиялық сапа факторын бағалау әдістемесі ұсынылған. Сандық осциллографпен өлшенетін шу сигналының амплитудаларының жылдам мәндерінің статистикасынан шу сапасының энтропиялық коэффициентін есептеу тәртiбi келтірілген.

Түйін сөздер: бүркемелейтін шуыл сигналы; шуыл генераторы; спектр талдағышы; энтропия коэффициенті.

**Н.А. Сейлова*, А.Б. Батыргалиев, Ж.А. Джангозин, Д.А. Байбатчаева,
Н. Нұрғабылов**

Казахский национальный исследовательский технический университет
имени К.И. Сатпаева, Алматы, Казахстан.

E-mail: *seilova_na@mail.ru*

МЕТОДИКА ОЦЕНКИ КАЧЕСТВА МАСКИРУЮЩИХ ШУМОВЫХ ПОМЕХ

Аннотация. В настоящее время увеличивается количество различных средств вычислительной техники, предназначенной для обработки, хранения и передачи информации.

Одной из таких угроз является наличие технических каналов утечки информации, возникающей вследствие информативных (опасных) побочных электромагнитных излучений технических средств ее обработки, хранения и передачи. Для защиты информации от утечек по каналам побочных электромагнитных излучений и наводок широко применяются системы пространственного электромагнитного зашумления. Однако не всегда осуществляется оценка качества маскирующих шумовых помех этих систем как в ходе их разработки, проектирования и производства, так и входе подтверждения их соответствия требованиям информационной безопасности. Как правило, ограничиваются проверкой соответствия диапазона рабочих частот, спектральной плотности, виду излучаемой помехи и др. В качестве показателя качества маскирующего шума, создаваемого генератором шума, используется энтропийный коэффициент качества шума.

Вопросам создания генераторов пространственного электромагнитного зашумления посвящено достаточное количество исследований. Но в настоящее время отсутствуют методы оценки качества шума, на основе которых имеются готовые технические решения. Таким образом, исследование характеристик качества генераторов пространственного электромагнитного зашумления является актуальным. В работе предложена методика оценки энтропийного коэффициента качества шума с использованием анализатора спектра и цифрового запоминающего осциллографа. Приведен порядок расчета энтропийный коэффициент качества шума по статистике мгновенных значений амплитуд шумового сигнала, измеренных цифровым осциллографом.

Ключевые слова: маскирующий шумовой сигнал; генератор шума; анализатор спектра; энтропийный коэффициент.

Introduction. Currently, electronics are used in almost all spheres of human life, ranging from simple means of communication to ensuring the security of the state as a whole. However, the use of electronic devices and computer technology carries a large number of information security threats. The basis for solving this

problem is the protection of information in the field of optics, electronics, radio engineering, acoustics and other sciences.

The main and at the same time one of the most dangerous technical channels of information leakage at informatization objects is the channel of side electromagnetic emissions (SEME). Such an information leakage channel is called electromagnetic (Khorev A.A. 2012:20, Vasiliev I.V. et.al, 2010:6, A.P. Zaitsev et al., 2017:442, Buzov G.A. 2017:586, ST RK GOST R 51275-2006, ST RK 1700-2007). Protection of electronic means of information processing and transmission (EM) from leakage through SERP channels is achieved by using passive and active protection methods. Passive protection methods include shielding, grounding, decoupling and filtering, and active methods include the use of systems of spatial electromagnetic noise and imitating (masking) interference (Zaitsev et al. 2017:442, ND TZI 3.3-001-07, ST RK 1698-2007).

The use of passive methods of protecting EM is the most preferable, since using them there are no problems associated with electromagnetic compatibility and the presence of unmasking signs of the operation of protective means. However, the use of passive methods for protecting EM is not always possible due to the complexity of their implementation, high cost, the need for additional development work, etc. In such cases, active protection methods are used, which lead to a decrease in the signal-to-noise ratio at the input of the receiving device of the reconnaissance means and, consequently, to a decrease in the controlled zone for the EM (Khorev A.A. 2012:20).

Currently, there are energy and non-energy methods of active protection. The non-energy (statistical) method of active masking consists in the emission of a special masking signal (interference) with a spectrum similar to the spectrum of informative SEME EM. The spectral density of masking interference should be higher than the spectral density of SEME, and its level should not exceed the levels of SEME (Zaitsev et al., 2017:442, ST RK 1698-2007). The difficulty in implementing a non-energy protection method lies in the need to use pulses of random amplitude, similar in shape and time-correlated radiation with informative SEME pulses. In this case, it is required to accurately determine all informative SEME for each EM sample and create an individual for this sample (or identical EM) masking noise simulator. As a rule, simulators or jammers are used to protect personal computers.

The essence of the energy method lies in the formation and emission into the surrounding space in the immediate vicinity of the operating EM of a masking broadband noise signal ("white noise") in the entire frequency range of informative SEME with a spectral level exceeding the levels of these SEME. Electromagnetic noise (electromagnetic interference, radio noise, radio interference, active masking interference) is understood as a time-varying electromagnetic phenomenon that does not contain information and can be superimposed on or combined with a useful signal. In the context of this work, they are intended to degrade or distort the normal operation of enemy electronic equipment. Active masking interference

creates a background at the input of the enemy receiver, which makes it difficult to detect informative SEME, their recognition and determination of parameters (ISO/IEC 19762-4-2011 2012:27, GOST R 55055-2012 2014:11). Therefore the use of generators of spatial electromagnetic noise (NG) should prevent or lead to the impossibility of intercepting informative SEME for their subsequent analysis and restoration of the original information or a significant complication of this process.

In the present work, the solution of the following problems is provided:

- determination of the main characteristics of the NG;
- conducting a review of existing methods for assessing the quality of masking noise interference and identifying their disadvantages;
- development of its own method for assessing the quality of masking noise interference;
- description of methods for measuring masking noise interference;
- determination of the quality of masking noise interference of some NG;
- search for the correlation of noise signals and the use of statistical methods (tests) for randomness as alternative methods for estimating the quality of masking noise interference.

Research Material and methods. Currently, there is a large number of NG with different technical characteristics and types of execution on the market (Zaitsev et al. 2017:442, Buzov G.A. 2017:586). Noise generators are available as a separate device or as a PCI board for a personal computer. NGs usually consist of a wideband signal generator and one or more antennas.

One of the most important requirements for the NG is the broadbandness of the noise signal spectrum and high uniformity of the noise power spectral density. For this reason, three schemes for generating a broadband noise signal are mainly used in noise generators (Khorev A.A. 2012:20):

1) classical method of forming direct noise interference. In this case, it is possible to use several noise sources operating in different frequency ranges. Noise resistors, diodes, transistors, zener diodes and other elements that form noise close in its characteristics to "white noise" can be used as primary sources of noise in such NGs.;

2) the use of a digital noise generator, the "digital" noise of which is a temporary random process, close in its properties to the process of physical noise and called the "pseudo-random process". Such generators form chaotic (pseudo-random) sequences of binary symbols and convert them into sequences of rectangular pulses of pseudo-random duration with pseudo-random intervals between them. Noise sources in such NGs can be microstrip elements, various integrated circuits, digital signal processors, programmable logic integrated circuits, and other elements (Zhanabayev Z.Zh. et al., 2010-7, Kogai G.D. et al., 2014-3, Gubanov D. et al., 1999:5, Dmitriev A.S. et al., 2009:16);

3) the use of a stochastic or chaotic method for generating a noise signal (Zhanabayev Z.Zh. et al., 2010-7, Ekhandel R. et al., 2014:4). The signal from the harmonic signal generator is fed to a power amplifier operating in a nonlinear

mode and loaded onto a non-autonomous nonlinear dynamic system in the form of a parallel nonlinear oscillatory circuit, in which the amplified signal is converted into noise stochastic interference.

However, noisy informative SEME can be filtered, and in case of poor quality masking, the enemy can gain access to the protected information (Gavrilov, I.V. et al., 2015:11).

In NG, to generate a broadband noise signal, there can be used a scheme for dividing the entire frequency range into subbands using a frequency multiplier (Khorev A.A. 2012:20). In such cases, the generated noise in different subbands will be correlated, i.e. have the same parameters except for the frequency.

This will make it possible to subtract noise in different ranges, in which informative SERP have a large amplitude (power) and further restore the protected information. It should be noted that the presence of additional factors in the form of the repetition of an informative signal, the level of its amplitude (power), etc. is also important (Batyrgaliyev A.B. 2019:3).

In addition, in the absence of complete randomness of the generated noise, there are used statistical methods of analysis, through which it is possible to identify the patterns of noise formation, including their periodicity. In this regard, an important problem arises related to assessing the quality of the noise signal generated by the NG.

In order to determine the estimated characteristics of the masking noise, information (non-energy or statistical) and energy methods are used. Information methods consider the statistical parameters of noise signals in the time domain and allow to directly determine the numerical noise quality factor. Based on the calculation of the mathematical expectation, dispersion and entropy of the instantaneous values of time readings and their envelope, the degree of approximation to some reference distributions is calculated. They are aimed at finding the degree of uncertainty of the instantaneous values of noise signals, expressed, for example, in terms of the entropy quality factor of the masking noise. When using this method of active masking, the NG emits a special masking signal (interference) with a spectrum similar to that of informative SERP. In this case, the spectral density of masking interference should be higher than the spectral density of SERP, and its level should not exceed the levels of SEME.

The energy method for information protection uses the postulate about the need to exceed the energy of noise over SEME in the entire frequency range. Therefore, in order to check the noise quality, integral indicators are used that take into account the excess of the noise level over the level of the informative signal (ST RK 1698-2007 Protection of information, Gavrilov, I.V. et al., 2015:11).

Result and discussion. One of the effective ways to protect computer equipment (CE) from information leakage through the channel of side electromagnetic emissions (SEME) is spatial electromagnetic noise (Khorev A.A. 2012:20).

For clarity, the application of the method for assessing the quality of the masking noises of the NG on Fig. 1 shows the histograms of the noise signal amplitudes

distribution. For the upper histogram, the entropy quality factor is 0.9, and for the lower one it is 0.7. The x-axis corresponds to the number of intervals, and the y-axis corresponds to the number of sample elements in the interval. The low entropy quality factor of the masking noise interference of the NG will not be able to ensure the security of the protected information.

A low entropy quality factor can, for example, be observed when the quality of the noise source is insufficient (noise diode, transistor, resistor, etc.) or in other cases. Such cases include the ability to control the level of the output signal of the NG, as well as the power supply of the NG with some boundary levels. For example, some manufacturers allow NG power supply within $220\text{ V} \pm 10\%$ at a power supply frequency of 50 Hz. This means that the power supply of the GS is possible within the range of 187-253 V. In some NGs, when they were powered at the boundary values, a significant deterioration in the entropy quality factor was observed. This circumstance directly increases the threat of leakage of information protected by the NG.

Entropy noise quality factor is used as an indicator of the quality of the masking noise K_m created by the NG, which characterizes the approximation of the noise power distribution law to the ideal “white” noise with a normal power distribution law.

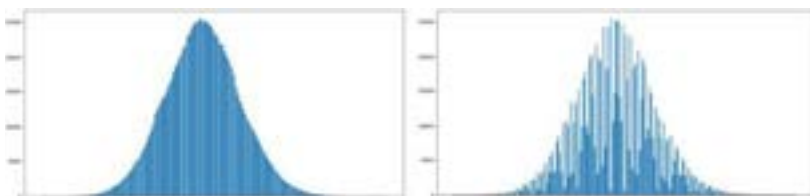


Figure 1. Histograms of noise signal amplitudes distribution: a) – $K_m = 0,9$; b) – $K_m = 0,7$

A technique for estimating ENQF using a spectrum analyzer and a digital storage oscilloscope is proposed.

The measuring installation for measuring and calculating the ENQF is shown in Figure 2.

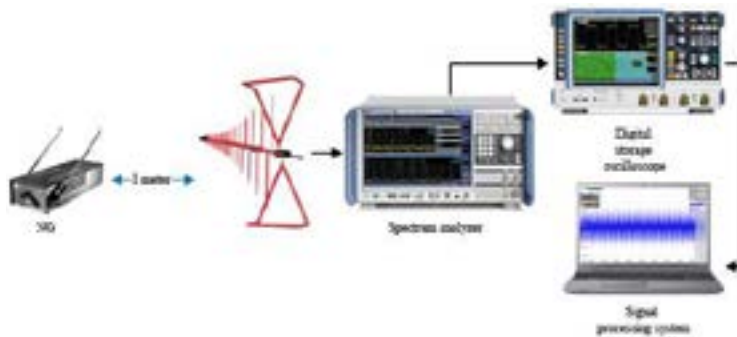


Figure 2. Measuring installation for measuring and calculating the ENQF

The measuring antenna is installed at a distance of 1 m from the noise generator antenna and connected to the antenna input of the spectrum analyzer. The maximum bandwidth (ΔF) of the spectrum analyzer is set. From the output of the intermediate frequency of the spectrum analyzer, the signal is fed to the input of a digital storage oscilloscope. The instantaneous values of the amplitude of the noise signal from the oscilloscope are transferred to the PC for processing.

The essence of the proposed methodology for assessing ENQF is following:

1) By analyzing the spectrum of the noise signal generated by the NG, the frequency intervals of the spectrum are selected, in which the greatest unevenness of the frequency response of the noise is observed. If the noise generator uses several channels for generating noise interference, then at least one such interval must be selected in each of them.

2) The spectrum analyzer is tuned sequentially to the center frequency of each of the frequency intervals. The electromagnetic noise signals received by the spectrum analyzer are converted into electrical signals and transmitted via an intermediate frequency to an oscilloscope that acts as an analog-to-digital converter.

3) Instantaneous values of the amplitude of the noise signal in *.csv format from the oscilloscope are sent to the signal processing system (personal computer with the calculation program) for further processing and calculation of the ENQF.

Below is the procedure for calculating the ENQF according to the statistics of the instantaneous values of the amplitudes of the noise signal generated by the NG:

1. The statistics of instantaneous values of noise signal amplitudes (n) with a volume of at least 106 elements is collected.

2. According to collected statistics $X = \{x_1, x_2, \dots, x_n\}$ a statistical series $\{x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(k)} \leq \dots \leq x_{(n)}\}$ is being constructed and calculated: average value (\bar{X}), dispersion (σ^2) and standard deviation (σ) according to the formulas:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \quad (2)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2} \quad (3)$$

3. The values of the statistical series $x(k)$ are grouped by selected non-overlapping intervals $(x_{(j-1)}; x_j)$, $j = 1, 2, \dots, m$, where m is the number of obtained intervals, and x_j are the upper boundaries of the intervals. When choosing the interval width, it is recommended to use the rule:

$$\Delta \leq \frac{(x_{(n)} - x_{(1)})}{2^r} \quad (4)$$

where Δ – maximum spacing width;

$x(1)$ и $x(n)$ – minimum and maximum elements of the statistical series, respectively;

r – bit depth of the used analog-to-digital converter of the measuring instrument.

The interval width Δ_j is equal to:

$$\Delta_j = x_j - x_{j-1}, \quad (5)$$

where $j = 1, 2, \dots, m$.

It is recommended that all intervals be chosen equal in width.

4. After choosing the intervals Δ_j for the sample $X = \{x_1, x_2, \dots, x_n\}$, the number n_j^* of the sample values $x_{(i)}$ that fall into the corresponding intervals is calculated. Based on the obtained values of n_j^* , the corresponding relative frequencies p_j^* and relative densities of sample values in each interval σ_j^* are calculated:

$$p_j^* = \frac{n_j^*}{n}. \quad (6)$$

$$\sigma_j^* = \frac{p_j^*}{h_j}. \quad (7)$$

The sum of the relative frequencies (p_j^*) must be equal to one, i.e:

$$\sum_{j=1}^m p_j^* = 1. \quad (8)$$

5. For cases when in any of the intervals n_j^* turns out to be equal to 0, you should combine this interval with the interval $(j-1)$ or $(j+1)$, recalculating the relative frequencies and relative densities in the newly formed intervals, or change Δ so, so that with a new partition, each of the intervals includes at least one sample value (x_j).

6. Based on the obtained data, Table 1 is compiled, which indicates the number of the interval (digit) j , the boundaries of the discharge $x_{j-1} - x_j$, the number of discharge n_j^* , the relative frequencies p_j^* and the relative densities σ_j^* of sample values.

Based on this table, a histogram of the distribution of instantaneous voltage values of the noise signal is constructed.

7. For each digit of the histogram, the entropy (H_j) is calculated according to the formula:

$$H_j = p_j^* \cdot \ln \sigma_j^* \quad (9)$$

Table 1 - Initial data for calculating ENQF

The number of the interval (discharge) j	1	2	...	m
Discharge $[x_{j-1}; x_j]$ boundaries	$[x_0; x_1]$	$[x_1; x_2]$...	$[x_{m-1}; x_m]$
Value n_j^*	n_1^*	n_2^*	...	n_m^*
Relative frequencies p_j^*	p_1^*	p_2^*	...	p_m^*
Relative densities σ_j^*	σ_1^*	σ_2^*	...	σ_m^*

The number of the interval (discharge) j	1	2	...	m
Entropy of the discharge H_j	H_1	H_2	...	H_m
Entropy of the noise signal H				

8. Next, the entropy of the noise signal (H) is calculated by the formula (10), the entropy power of the noise signal (P_3) by the formula (11) and the entropy quality factor of the instantaneous values of masking noise voltages (K) by the formula (12):

$$H = - \sum_{j=1}^m H_j. \tag{10}$$

$$P_3 = \frac{e^{2H}}{2\pi e} \tag{11}$$

$$K_{\text{ш}} = \frac{P_3}{\sigma^2}. \tag{12}$$

The calculated value of $K_{\text{ш}}$ is compared with the normalized value of $K_{\text{шн}}$ set for the given type of noise generator.

The proposed method has been tested in measuring the ENQF of noise generators installed at informatization facilities.

The tests were carried out using a laboratory complex consisting of an AI-5.0 active measuring antenna, an R&S FSW8 digital spectrum analyzer, an R&S RTO 1022 digital storage oscilloscope, and a laptop-based signal processing complex.

As an example, Figures 3 and 4 show the masking noise spectra generated by the noise generators LNG- 503, Gnome-3, Salyut 2000 B and Sonata-R2, and Table 1 shows the measured values of the entropy noise quality factor of these generators obtained during the research.

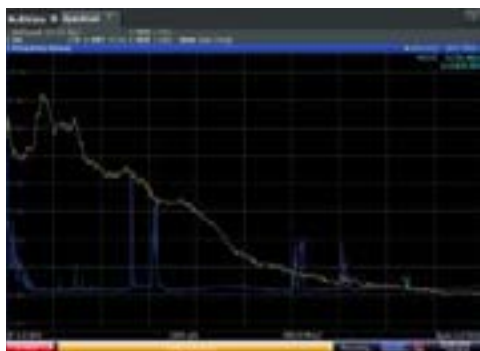


Figure 3.1



Figure 3.2

Figure 3. Spectra of masking noises generated by noise generators "LNG-503" (3.1) and "Gnom-3" (3.2)

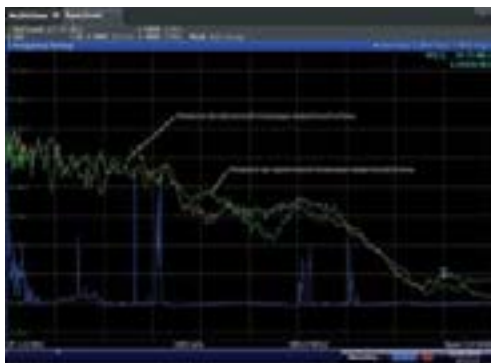


Figure 4.1

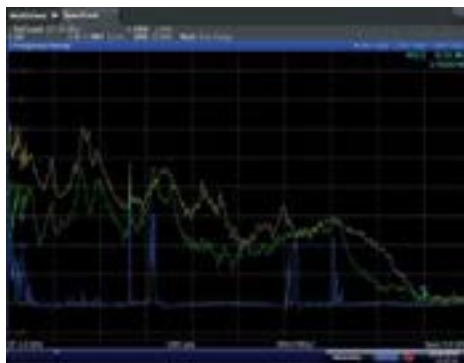


Figure 4.2

Figure 4. Spectra of masking noises generated by noise generators "Salyut 2000 B" (4.1) and "Sonata-R2" (4.2)

As can be seen from Table 2, the measured values of the entropy noise quality factors generated by noise generators basically correspond to their passport values.

Table 2 - Measured values entropy noise quality factor of noise generators

Noise generator type	Spectrum analyzer tuning frequency, MHz	Spectrum analyzer bandwidth, MHz	Radiation power	Measured value ENQF
"LNG-503"	500	80	Maximum	0,90
"Gnome-3M"	500	80	Maximum	0,97
"Salyut-2000B"	500	80	Maximum	0,96
"Sonata – R2"	500	80	Maximum	0,88
			Maximum	0,98

Thus, the proposed technique makes it possible to measure the entropy noise quality factors generated by noise generators of spatial electromagnetic noise systems without galvanic connection of measuring instruments to generators and provides sufficient measurement accuracy.

Carrying out manual calculations using the proposed alternative method for estimating the quality of masking noise interference is a very laborious and lengthy process. This is due to the need to process a large number of values (at least 1 million elements).

In this regard, there has been developed a software to implement the automatic calculation of measurement results. The program consists of several functions and the calculation itself.

Figure 5 shows an example of the calculation program result.

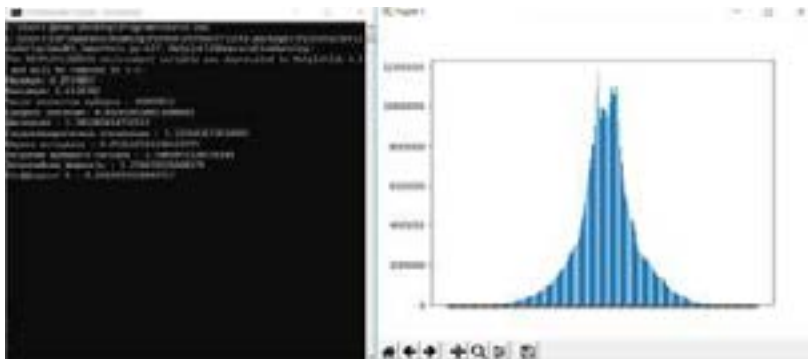


Figure 5. An example of the calculation program result

Conclusions. Spectrum analyzers (or other measuring receivers), digital storage oscilloscopes, or mixed signal oscilloscopes are considered appropriate to make the necessary measurements.

In this case, you should pay attention to the parameters of the selected measuring instruments: operating frequency range and bandwidth.

If you select an oscilloscope with the required bandwidth of 2-3 GHz (operating frequency range of the NG), then for measurements on the air this will be a practically impossible task. This is because even the best performing premium spectrum analyzers have less bandwidth. For example, the Keysight UXA (N9040B) spectrum analyzers have a maximum bandwidth of 1 GHz (real-time 510 MHz) and the R&S®FSW - 800 MHz (expandable to 5 GHz with an RTO2064 oscilloscope) [19, 20]. However, the cost of such devices is very high.

In this regard, to carry out measurements on the air using a spectrum analyzer, the operating frequency range of the NG will need to be divided into equal subbands that fit into the bandwidth of the spectrum analyzer.

Acknowledgements. *This research has is funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan under the theme “Quality assessment of spatial electromagnetic noise in active information protection systems” (Grant №AP08856630).*

Information about authors:

Seilova Nurgul – Candidate of technical science, Professor-Researcher Kazakh National Research Technical University named after K.I.Satpayev, Dean faculty of Computer Technologies and Cybersecurity, IITU, Almaty, Kazakhstan; H-index: 2 (Scopus), Scopus ID: 57200070158 <https://www.scopus.com/authid/detail.uri?authorId=57200070158>, <https://orcid.org/0000-0003-3827-179X>;

Batyrgaliyev Askhat – PhD, Department of "Cybersecurity, Information Processing and Storage" Kazakh National Research Technical University named after K. I. Satpayev, Almaty, Kazakhstan; Scopus ID: 57219016615, <https://www.>

scopus.com/authid/detail.uri?authorId=57219016615, <https://orcid.org/0000-0002-1103-8659>;

Zhanibek Jangozin – Senior researcher, Department of "Cybersecurity, information processing and storage" Kazakh National Research Technical University named after K.I. Satpayev, Almaty, Kazakhstan, ORCID: *<https://orcid.org/0000-0003-0047-6124>*;

Dana Baibatchayeva – Senior researcher, Department of "Cybersecurity, information processing and storage" Kazakh National Research Technical University named after K.I. Satpayev, Almaty, Kazakhstan, ORCID: *<https://orcid.org/0000-0001-7287-4864>*;

Nurgabilov Nurbol – Researcher Department of "Cybersecurity, Information Processing and Storage" Kazakh National Research Technical University named after K.I. Satpayev, 22, Almaty, Kazakhstan, *<https://orcid.org/0000-0002-5334-7003>*.

REFERENCES:

Batyrgaliyev A.B. Approaches to noise quality estimation of spatial electromagnetic noise generators // Contemporary technologies in science and education - STNO-2019 [text]: collection of works of the II international scientific and technical forum: in 10 volumes, Volume 5/ edited by O.V. Mylovzorova. - Ryazan: Ryazan State Radio Technical University, 2019; Ryazan – pp. 86: ill. - pp. 17-19.

Buzov G.A. Protecting restricted access information from leakage through technical channels. - M.: Hotline - Telecom, 2017. – pp. 586, ill.

Dmitriev A.S., Efremova E.V., Nikyshov A.Yu., Panas A.I. Chaos generators: from vacuum devices to nanocircuits // Radioelectronics, nanosystems, information technologies. - 2009. Vol. 1. - No. 1-2. - pp. 6-22.

Ekhande R., Deshmukh S, "Chaotic Signal for Signal Masking in Digital Communications", IOSR Journal of Engineering Vol. 04, Issue 02 (February, 2014), V5, pp 29-33.

Gavrilov, I.V. Methods of estimating the quality of masking noise // Proceedings of the St. Petersburg Institute of Informatics and Automation of the Russian Academy of Sciences. - 2015. Issue 6 (43). - pp.179-190.

GOST R 55055-2012 Industrial radio interference. Terms and definitions. - M.: Standartinform, 2014. – pp.11.

Gubanov D., Dmitriev A., Panas A., Starkov S., Steshenko V. Chaos generators in integral execution // Chipnews. News about microcircuits. - 1999. - No. 8. - pp. 9-14.

ISO/IEC 19762-4-2011 Information technologies. Technologies of automatic identification and data collection (AIDC). Harmonized dictionary Part 4. General terms in the field of radio communication. - M.: Standartinform, 2012. – pp. 27.

Khorev A.A. Ways to protect informatization objects from information leakage through technical channels: spatial electromagnetic noise// Special Technique. - M.: 2012. - No. 6 - pp. 37-57.

Kogai G.D., Ten T.L. Methods and models of chaotic processes in communication systems // Modern science-intensive technologies. - 2014. - No. 10. - pp. 61-64.

N9040B UXA signal analyzer [Electronic resource]. – <https://www.keysight.com/ru/pdx-x202152-pn-N9040B/uxa-signal-analyzer-multi-touch-3-hz-to-50-ghz?pm=spc&nid=-32508.1150334&cc=RU&lc=rus>.

ND TZI 3.3-001-07 Protection of information at the objects of information activity. Creation of a complex of technical protection of information. Procedure for development and implementation of information protection measures.

R&S®FSW spectrum and signal analyzer [Electronic resource]. – https://www.rohde-schwarz.com/ru/product/fsw-productstartpage_63493-11793.html.

ST RK GOST R 51275-2006 Information security. Informatization object. Factors affecting information. General provisions.

ST RK 1700-2007 Technical protection of information in office premises. General technical requirements.

ST RK 1698-2007 Information security. Protection of information from technical intelligence and from its leakage through technical channels at the facility of computer facilities. Protection methods.

ST RK 1698-2007 Protection of information. Protection of information from technical intelligence and its leakage through technical channels at the facility of computing equipment. Protection methods.

Technical means and methods of information protection. Textbook for universities / A.P. Zaitsev, A.A. Shelupanov, R.V. Meshcheryakov. Ed. by A.P. Zaitsev and A.A. Shelupanov. - 7th ed., Rev. - M.: Hotline - Telecom, 2017. – pp. 442: ill

Vasiliev I.V., Kozin, I.D., Fedulina, I.N. Data protection. Tutorial. - Almaty: AUPET, 2010. – pp. 76.

Zhanabayev Z.Zh., Almasbekov N.E., Imanbayeva A.K., Manapbayeva A.B., Akhtanov S.N. Information protection by dynamic chaos with phase control. // Materials of the 7th international scientific conference. "Chaos and structure in nonlinear systems", June 15-17, 2010. Karaganda. - pp. 13-20.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 134-145
<https://doi.org/10.32014/2022.2518-1726.162>

УДК 004.942

**А.Ш. Хасенов*, М.А. Кантуреева, Д.А. Тусупов, А.С. Омарбекова,
Г.Б. Абдикеримова**

Евразийский национальный университет им. Л.Н. Гумилева,
Астана, Казахстан.

E-mail: khssnv04@gmail.com

ПОДХОД К РЕАЛИЗАЦИИ МОДЕЛИ ЭВАКУАЦИИ В СИСТЕМЕ АГЕНТНОГО МОДЕЛИРОВАНИЯ

Аннотация. В последнее время нападение на учебные заведения стало серьезной угрозой безопасности международного сообщества. За последнее десятилетие часто совершались террористические атаки по всему миру. В местах массового скопления людей нет идеального плана эвакуации. В статье исследуется важность моделирования эвакуационных выходов в учебных заведениях с помощью клеточного автомата, рассматривается исследование имитационной модели эвакуации, а также приводятся общие принципы работы систем агентского моделирования и средств отображения смоделированных данных, описывается структура системы, предлагается модель эвакуации. Также решение проблемы управления движением объектов обсуждено путем составления плана эвакуации или управления движением с помощью оператора. Важна роль модели клеточного автомата в исследованиях эвакуации толпы. Эвакуация людей с реальной ситуацией очень важна для создания имитационной модели. Рассмотрено два типа исследования имитационной модели эвакуации: макромодель и микромодель. Рассмотрено клеточный массив, в котором нет клеток с одинаковыми именами. Также рассмотрено два основных режима применения глобального оператора: синхронный и асинхронный. Синхронный режим – это состояние клеток-соседей на текущей итерации. А при асинхронном режиме каждая клетка вычисляет функцию перехода состояний соседей и меняет свое состояние. Исследована FF-модель, которая относится к классу дискретных стохастических моделей с индивидуальным представлением людей.

Ключевые слова: клеточный автомат, скулшутинг, террористическая атака, эвакуация, управления движением людей, множество клеток, оператор, клетка массива.

Финансирование: Работа поддержана Комитетом науки МВОН РК, грант № AP08855497.

**А.Ш. Хасенов*, М.А. Кантуреева, Д.А. Тусупов, А.С. Омарбекова,
Г.Б. Абдикеримова**

Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан.
E-mail: khssnv04@gmail.com

АГЕНТТІК МОДЕЛЬДЕУ ЖҮЙЕСІНДЕ ЭВАКУАЦИЯ МОДЕЛІН ЖҮЗЕГЕ АСЫРУ ТӘСІЛІ

Аннотация. Соңғы уақыттардағы оқу орындарына жасалған шабуыл халықаралық қауымдастықтың қауіпсіздігіне үлкен қауіп төндірді. Соңғы онжылдықта бүкіл әлемде лаңкестік шабуылдар жиі болды. Адамдар көп жиналатын жерлерде эвакуацияның ұтымды жоспары жоқ. Мақалада клеткалық автомат арқылы оқу орындарындағы эвакуациялық шығуларды модельдеудің маңыздылығы зерттеліп, эвакуацияның имитациялық моделін зерттеу қарастырылады, сонымен қатар агенттік модельдеу жүйелері мен модельденген деректерді көрсету құралдарының жұмысының жалпы принциптері келтірілген, жүйенің құрылымы сипатталған, эвакуация моделі ұсынылады. Сондай-ақ, объектілердің қозғалысын басқару мәселесін шешу эвакуациялау жоспарын құру немесе оператордың көмегімен қозғалысты басқару арқылы талқыланады. Топтық эвакуациялауды зерттеудегі клеткалық автомат моделінің рөлі маңызды. Имитациялық модель құруға шынайы болған адамдардың эвакуациялау оқиғасы өте маңызды. Эвакуацияның имитациялық моделін зерттеудің екі түрі қарастырылған: макромоделі және микромоделі. Бірдей атаулары жоқ клеткалық автомат массиві талданды. Сондай-ақ, жаһандық операторды қолданудың екі негізгі режимі қарастырылды: синхронды және асинхронды. Синхронды режим – қазіргі итерациядағы көрші торлардың жағдайы. Асинхронды режимде – әр ұяшық көршілердің күйін ауыстыру функциясын есептейді және өзінің күйін өзгертеді. Адамдарды жеке сипаттайтын дискретті стохастикалық модельдер класына жататын FF моделі зерттелді.

Түйін сөздер: клетка автоматы, скулшутинг, террористік шабуыл, эвакуация, адамдардың қозғалысын басқару, көптеген клеткалар, оператор, массивтің клеткасы.

Қаржыландыру: Жұмысты ҚР ЖБФМ Ғылым комитеті, №AP08855497 гранты қолдады.

**A. Khassenov*, M. Kantureyeva, D. Tussupov, A. Omarbekova,
G. Abdikerimova**

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.
E-mail: *khssnv04@gmail.com*

APPROACH TO THE IMPLEMENTATION OF EVACUATION MODEL IN THE AGENT-BASED MODELING SYSTEM

Abstract. Recently, the attack on educational institutions has become a serious threat to the security of the international community. There have been frequent terrorist attacks around the world over the past decade. There is no ideal evacuation plan in crowded places. The article examines the importance of modeling evacuation exits in educational institutions using a cellular automaton, examines the study of an evacuation simulation model, and also provides general principles of operation of agency modeling systems and means of displaying simulated data, describes the structure of the system, and suggests an evacuation model. Also, the solution to the problem of controlling the movement of objects is discussed by drawing up an evacuation plan or motion control with the help of an operator. The role of the cellular automaton model in crowd evacuation studies is important. Evacuation of people with a real situation is very important for creating a simulation model. Two types of research of the simulated evacuation model are considered: a macro-model and a micromodel. A cellular array is considered in which there are no cells with the same names. Two main modes of application of the global operator are also considered: synchronous and asynchronous. Synchronous mode is the state of neighboring cells at the current iteration. And in asynchronous mode, each cell calculates the state transition function of its neighbors and changes its state. The FF-model, which belongs to the class of discrete stochastic models with an individual representation of people, is investigated.

Key words: cellular automaton, schoolshooting, terrorist attack, evacuation, traffic control of people, set of cells, operator, array cell.

Funding: *The work was supported by the Scientific Committee of the MHES of the RK, grant No. AP08855497.*

Введение. Различные чрезвычайные ситуации приводят к возникновению стихийных бедствий, аварий, катастроф с многочисленными человеческими жертвами, огромными материальными потерями. Разрушительные последствия чрезвычайных ситуаций разнообразны: механические, термические, химические, радиационные, биологические, психологические и т.д.

С помощью устойчивости населенных пунктов реагирование на чрезвычайные ситуации является главной задачей государства и проблемой нашего времени. А что касается помощи пострадавшим во время надлежащих

операций, то ответственность населения заключается в том, чтобы уменьшить потери и максимально помочь людям в сложной ситуации.

Использование огнестрельного оружия, ножей, топоров и огнестрельного оружия в учебных заведениях, которые сегодня занимают много места, является глобальной эпидемией. В соседних странах о школьной стрельбе или скулшутинге (от англ. School - школа и стрелять - стрелять) начали рассматривать как отдельную тему.

В последнее время нападение на учебные заведения стало серьезной угрозой безопасности международного сообщества. За последнее десятилетие террористические атаки по всему миру часто совершались с использованием различного оружия, таких как нож, топор и т.д. Вооруженные атаки сложно осуществить, поэтому террористы обычно имеют хорошо продуманные планы. Террористы вызвали массовую панику и тревогу в нашем сообществе, о чем свидетельствует большое число жертв. Скулшутинг следует отличать от террористических актов (школа в Беслане). В местах массового скопления людей не существует идеального плана ликвидации или эвакуации, когда речь идет о вооруженном нападении. В некоторых общественных местах нет даже запасного выхода для эвакуации, особенно в странах СНГ, где все мы видели через средства массовой информации, как люди пытаются прыгнуть со второго этажа. И в таких ситуациях невозможно быстро эвакуироваться, что приведет к большому количеству жертв (Hirsch и др., 2015; Brandrud и др., 2017).

В основном многие исследования по вооруженному нападению обращают внимание только на предотвращение вооруженного нападения или уничтожение террористов. И исследования по эвакуации людей во время вооруженных нападений рассматриваются с недостаточной информацией. В таких исследованиях нужно учесть много недостатков и пробелов. И, во-первых, при вооруженных нападениях пострадавшие люди не могут сразу эвакуироваться, так как не имеют соответствующего опыта. В таких стрессовых ситуациях жертва может попасть в шоковое состояние, что замедляет эвакуацию. Во-вторых, во время нападения люди не могут сразу понять, что происходит и что нужно делать, так как нет исследования и соответствующего плана эвакуации. Таким образом, это привело к большому количеству жертв в результате вооруженных нападений. В-третьих, во многих общественных местах нет эвакуационных проходов или неправильно применяются. В-четвертых, важным фактором является то, что при скоплении людей в эвакуационных проходах пострадавшие часто страдают от столкновений, вызванных эмоциями (давка и гибель людей).

Критерии успешной эвакуации различаются в разных ситуациях. Например, во время стихийных бедствий и вооруженных нападений стандарты эвакуации различаются. Бедствия, такие как пожары и землетрясения, цель человека – эвакуироваться из здания или держаться подальше (Ding и др., 2017). А когда происходит вооруженное нападение, целью человека является держаться

подальше от террориста. По всему миру в больших городах построены крупные учебные заведения, в которых может поместиться много студентов и школьников. Причем в таких зданиях эвакуация идет очень медленно, во время чрезвычайной ситуации.

Сейчас такие задачи, как управление трафиком рассматриваются с помощью планов или операторов. Для начала до людей доводят маршруты до выхода с этажа, которые являются кратчайшими по расстоянию. Во втором варианте оператор получает информацию от системы наблюдения и на основании этих данных решает о направлениях движения. Оба подхода являются неприемлемыми при управлении движением людей в среде. Например, затруднительно предусмотреть все возможные места возникновения возгорания в здании. Во многих зданиях (торговые центры, торгово-развлекательные комплексы и т.д.) может развиваться необходимость изменения направления движения при чрезвычайных ситуациях.

Обсуждение. Роль модели клеточного автомата в исследованиях эвакуации толпы. В последние годы от нападения на учебное заведение люди пострадали от злоумышленников, а также от столпотворения, вызванной паникой. Поэтому эвакуация толпы в чрезвычайных ситуациях является ключевым аспектом при проектировании здания с аварийным выходом. Клеточные автоматы используются для моделирования динамических объектов. По сути, во многих исследованиях использовались экспериментальные методы и методы моделирования для изучения поведения толпы во время эвакуации. И, как обычно, после реального события используются экспериментальные методы с помощью анкеты, после чего происходит информационная задержка. К тому же, такие эксперименты очень дороги и их сложно смоделировать по разным сценариям. Поэтому эвакуация людей с реальной ситуацией очень важна для создания имитационной модели и широко используется (Kim и др., 2018; Zheng и др., 2019). Метод имитационной модели включает в себя проектирование обычных строительных объектов и других атрибутов толпы. А также можно рассмотреть отдельного человека в качестве объекта исследования и смоделировать процесс эвакуации. Такая модель поможет оценить эффективность управления объектом и поддержки при принятии решений в чрезвычайных ситуациях.

Существует два типа исследования имитационной модели эвакуации: макромодель и микромодель. Модель макростимуляции трансформируется в задачу выбора пути эвакуации с использованием метода теории графов. И обычно используются такие методы, как максимальный поток с низкой стоимостью, метод очереди и кратчайший путь (Liu и др., 2018; Shin и др., 2019). А микромодели принимают за объекты характеристики индивидуального движения, поведения человека при эвакуации.

Модель обычного типа включает в себя модель визуальных препятствий, клеточную автоматизированную модель, решеточный газ и социальную силу (Jiang и др., 2020) (Guo и др., 2012). Модель клеточного автомата (КА) может

охватывать такие аспекты, как сложная вычислительная функция и различные динамические характеристики с небольшими вычислениями. И это ключевой фактор, используемый в исследованиях эвакуации толпы.

Как говорил Franovetter (Franovetter и др., 1978), эмоции играют важную роль в таких критических ситуациях, что человек может их воспринимать и распространять, и предложил эмоциональный порог. В то время McCrae (McCrae и др., 1996) предложил модель личности ОКЕАН с пятью векторными измерениями. А также Helbing (Helbing и др., 2000) подтвердил, что эмоции людей при побеге заразительны. И Wang (Wang и др., 2016) пришел к выводу, что эмоциональное заражение распространяется на поведение отдельных лиц, групп.

Методы и материалы. Предварительные сведения из теории клеточных автоматов. Приведем некоторые сведения из теории КА о формальном представлении КА (Бандман и др., 2006).

Клеточный автомат обычно определяется множеством клеток, плотно заполняющих дискретное D -мерное пространство.

Клеткой называется пара $(x, m) \in A \times M$, где $x \in A$ – символ алфавита A , обозначающий состояние клетки, $m \in M$ – имя клетки из множества имён M , обозначающие местоположение клетки в дискретном пространстве. Алфавит может быть булевым, $A_B = \{0;1\}$, вещественным, $A_R = [0;1]$, и символьным, $A_S = \{a, b, c, d, \dots\}$. Состояние могут быть заданы переменными, определёнными на одном из этих алфавитов.

Множество клеток $\Omega = \{(x, m)\}$, в котором нет клеток с одинаковыми именами, называется клеточным массивом. Клеточные массивы, в которых состояния определены на множестве A , а имена на множестве M , составляют класс клеточных массивов $\Omega = \{(A, M)\}$. На множестве имён определены именующие функции $\phi: M \rightarrow M$, которые определяют пары взаимодействующих клеток.

Именующая функция $\phi(m)$ определяет для любой клетки m одного из её соседей. Для декартовых клеточных массивов именующие функции имеют вид сдвига, $\phi(i, j) = (i + a, j + b)$, a и b – константы, которые определяют пары взаимодействующих клеток.

Конечное множество именующих функций называется шаблоном соседства: $T(m) = \{\phi_0(m), \phi_1(m), \dots, \phi_q(m)\}$.

Для каждой клетки $m \in M$ шаблон $T(m)$ определяет множество её соседей. Нумерация клеток в $T(m)$ может быть любой. Принято считать, что $\phi_0(m) = m$.

Выражение вида $Conf(m) = \{(x_0, \phi_0(m)), (x_1, \phi_1(m)), \dots, (x_q, \phi_q(m))\}$ называется локальной конфигурацией, если существует такое r , что расстояние от m до любого $\phi_k(m) \in T(m)$, $d(m, \phi_k(m)) < r$, где $T(m)$ – множество входящих в $Conf(m)$ именующих функций, называемое определяющим шаблоном локальной конфигурации.

Параллельной подстановкой (далее подстановкой) называется выражение вида

$$\theta(m): Conf(m) * Conf''(m) \rightarrow Conf'(m), \text{ где}$$

$$Conf(m) = \{(x_0, \phi_0(m)), (x_1, \phi_1(m)), \dots, (x_q, \phi_q(m))\},$$

$$Conf'(m) = \{(h_0, \phi_0(m)), (h_1, \phi_1(m)), \dots, (h_q, \phi_q(m))\},$$

$$Conf''(m) = \{(g_0, \psi_0(m)), (g_1, \psi_1(m)), \dots, (g_n, \psi_n(m))\}.$$

$Conf(m)$ называется базовой конфигурацией (или базой), $Conf''(m)$ – контекстом, $Conf'(m)$ – правой частью подстановки. База и правая часть подстановки имеют один тот же определяющий шаблон. Шаблон контекста от них отличается. Символы h_k в клетках правой части обозначают функции от значений состояний в клетках базы и контекста:

$$h_k = H(x_0, x_1, \dots, x_q, g_0, \dots, g_n).$$

Они называются функциями переходов и могут быть булевыми, вещественными, а также простыми заменами символов в соответствии с используемым алфавитом.

Параллельная подстановка является локальным оператором КА-модели. Результат применения локального оператора ко всем клеткам массива Ω называется глобальным оператором $\theta(\Omega)$.

Существует два основных режима применения глобального оператора вычисления следующего глобального состояния $\Omega(t+1)$: синхронный и асинхронный. Синхронный режим предполагает, что аргументы функции переходов – это состояние клеток-соседей на текущей итерации t . На каждой итерации клетки вычисляют значения нового состояния и затем все клетки синхронно заменяют старые состояния на новые. При асинхронном режиме каждая клетка вычисляет функцию перехода тех значений состояний соседей, которые на данный момент имеют место, и сразу меняет свое состояние.

Итеративный процесс переходов из состояния в состояние называется эволюцией КА.

Границы областей моделирования представляются особыми клетками, которые имеют граничные правила переходов.

Постановка задачи. Геометрия области моделирования представляется в виде плоской области, разбитой на ячейки. Каждая ячейка отражает часть рассматриваемой области размера $0,4 \times 0,4 \text{ м}^2$ (это средний размер, занимаемый человеком в толпе (Kirchner и др., 2002)). Ячейки разделяются на ячейки-частицы, ячейки-препятствия, ячейки-выходы и пустые.

Область моделирования удобно представлять в виде двух массивов: массив препятствий W и массив F^t . t – номер дискретного временного шага или номер итерации КА.

Массив препятствий W задается изначально и не изменяется со временем.
 $W = \{w_{ij} : i = \overline{1, I}, j = \overline{1, J}\}$, где

$$w_{ij} = \begin{cases} 1, & \text{ячейка}(i, j) \text{ занята стеной (препятствием);} \\ 2, & \text{в ячейке}(i, j) \text{ выход;} \\ 0, & \text{ячейка}(i, j) \text{ свободна.} \end{cases}$$

Массив частиц F^t изменяется со временем и определяет эволюцию КА.

$$F^t = \{f_{ij}^t : i = \overline{1, I}, j = \overline{1, J}\}, \text{ где}$$

$$f_{ij}^t = \begin{cases} 1, & \text{ячейка}(i, j) \text{ занята частицей и } w_{ij} = 0; \\ 0, & \text{ячейка}(i, j) \text{ свободна} \end{cases}$$

Таким образом, задали состояние клетки $f \in A$ и алфавит $A = \{0;1\}$. Множество имён $M = \{(i, j) : i = \overline{1, I}, j = \overline{1, J}\}$.

В момент времени $t=0$ количество частиц N расположено на массиве F^0 . Число N не должно превышать количества свободных от ячеек-препятствий мест. Каждая ячейка может быть занята только одной частицей, частицы могут располагаться на свободных от препятствий местах:

$$(w_{ij} = 1) \cap (f_{ij}^t = 1) = \emptyset, \quad i = \overline{1, I}, j = \overline{1, J}.$$

Полагается, что геометрия рассматриваемой области «известна» частицам и отражена в «карте местности» - статическом поле S . Поле S совпадает с рассматриваемой областью и также дискретизировано на ячейки $0,4 \times 0,4 \text{ м}^2$. $S = \{s_{ij} : i = \overline{1, I}, j = \overline{1, J}\}$. Ячейки поля содержат минимальное расстояние до ближайшего выхода.

На каждом шаге t частица может переместиться в одну из четырех свободных соседних ячеек или остаться на прежнем месте (окрестность фон Неймана), таблица 1.1. Шаблон соседства

$$T(i, j) = \{\phi_k(i, j) : k = \overline{0, 4}\} = \{(i, j), (i, j-1), (i-1, j), (i, j+1), (i+1, j)\}.$$

$\phi_k(i, j)$ – соответствующая соседняя клетка, которую также можно рассматривать как направление движения.

Таблица 1.1 – Шаблон соседства $T(i, j)$

	(i-1, j)	
(i, j-1)	(i, j)	(i, j+1)
	(i+1, j)	

Скорость частиц в модели $v_{\max} = 1$ ячейка.

Направление движения частицы рассматривается как случайная величина. Пусть случайная величина ξ – направление движения частицы с координатами (i, j) на расчетном шаге t . $\rho_{\phi_k(i, j)}$ – вероятность переместиться в соседнюю ячейку $\phi_k(i, j)$. Тогда ξ – имеет распределение:

ξ	$\phi_0(i, j)$	$\phi_1(i, j)$	$\phi_2(i, j)$	$\phi_3(i, j)$	$\phi_4(i, j)$
ρ_ξ	$P \phi_0(i, j)$	$P \phi_1(i, j)$	$P \phi_2(i, j)$	$P \phi_3(i, j)$	$P \phi_4(i, j)$

Целью передвижения частиц является ближайший или заданный выход. Направление движения частицы определяется на основе вероятностей переходов в каждом направлении для каждой частицы в каждый дискретный шаг времени и набором правил переходов.

Вывод. Основным предметом исследований является определение способа вычисления вероятностей и определения правил перехода так, чтобы воспроизводить направленное движение частиц к выходу в зданиях и сооружениях любой геометрии с приближенным выполнением выделенных в первой главе свойств движения людей.

Для решения поставленной задачи за основу была взята FF-модель (Burstedde и др., 2001; Nishinari и др., 2004). Она относится к классу дискретных стохастических моделей с индивидуальным представлением людей. Физические параметры унифицируются: все частицы имеют одну и ту же площадь проекции и скорость.

Направление движения рассматривается как случайная величина и определяется на основе переходных вероятностей. Переход на новые позиции осуществляется посредством правил переходов согласно выбранным направлениям движения.

При построении модели использовались следующие идеи других учёных. Идея поля расстояний до выхода S и экспоненциальная форма переходных вероятностей была взята из FF модели.

При вычислении вероятностей переходов учитывается карта местности поле S , плотность частиц в направлении и близость стен в направлении. Идея использования этих компонент присутствует в том или ином виде и в других моделях движения людей, но именно взаимодействие параметров модели даёт реализацию введенных стратегий кратчайшего пути и быстрого пути и приводит к более адекватному моделированию динамики движения людей. Параметрами модели можно регулировать динамику движения в зависимости от пространственной ситуации.

Information about the authors

A. Khasenov – Doctoral student of the Department "Information Systems" of the Faculty of Information Technologies of the L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: khssnv04@gmail.com, <https://orcid.org/0000-0002-6190-6776>;

M. Kantureyeva – Acting Associate Professor of the Department "Information Systems" of the Faculty of Information Technologies of the L.N. Gumilyov

Eurasian National University, PhD. Astana, Kazakhstan, E-mail: ma_khantore@mail.ru, <https://orcid.org/0000-0001-5904-820X>;

D. Tussupov – Professor. Head of the Department "Information Systems" of the Faculty of Information Technologies of the L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: tussupov@mail.ru, <https://orcid.org/0000-0002-9179-0428>;

A. Omarbekova – Candidate of Technical Sciences, Director of the Department of Digital Development and Distance Learning, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: omarbekova_as@enu.kz, <https://orcid.org/0000-0002-9272-8829>;

G. Abdikerimova – PhD, Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: gulzira1981@mail.ru, <https://orcid.org/0000-0002-4953-0737>.

ЛИТЕРАТУРЫ

Бандман О.Л. Клеточно-автоматные модели пространственной динамики / О.Л. Бандман // Системная информатика: сб. научн. тр./ по ред. А.Г. Марчука. – Новосибирск: Издательство СО РАН, 2006. –С. 59-113.

A.S. Brandrud, Michael Bretthauer, Guttorm Bratteb, et al., Local emergency medical response after a terrorist attack in Norway: a qualitative study, *Bmj Qual. Saf.* 26 (10) (2017).

B. Liu, Z. Liu, D. Sun, and C. Bi, "An evacuation route model of crowd based on emotion and geodesic," *Mathematical Problems in Engineering*, vol. 2018, Article ID 6585102, 2018.

Burstedde C. Simulation of pedestrian dynamics using a two-dimensional cellular automaton / C. Burstedde, K. Klauack, A. Schadschneider, J. Zittartz // *Physica A.* – 2001. – Vol. 295. – P. 507-525.

D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487-490, 2000.

H. Kim and S. Han, "Crowd evacuation simulation using active route choice model based on human characteristics," *Simulation Modelling Practice and Theory*, vol. 87, pp. 369-378, 2018.

Kirchner A. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics / A. Kirchner, A. Schadschneider // *Physica A.* – 2002. – Vol. 312. – P. 260-276.

M. Franovetter, "Threshold models of collective behavior," *American Journal of Sociology*, vol. 83, no. 6, pp. 1420-1443, 1978.

M. Hirsch, Pierre Carli, Remy Nizard, et al., The medical response to multisite terrorist attacks in Paris, *Lancet* 386 (10012) (2015) 2535-2538.

N. Ding, T. Chen, H. Zhang, Experimental study of elevator loading and unloading time during evacuation in high-rise buildings, *Fire Technol.* 53 (1) (2017) 29-42.

N. Ding, H. Zhang, T. Chen, Simulation-based optimization of emergency evacuation strategy in ultra-high-rise buildings, *Natural Hazards* 89 (3) (2017) 1037-1184.

Nishinari K. Extended floor field CA model for evacuation dynamics / K. Nishinari, A. Kirchner, A. Namazi, A. Schadschneider // *IEICE Transactions on Information and Systems*, E87-D. – 2004. – P. 726-732.

R.R. McCrae, P.T. Costa and J.S. Wiggins, *Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-Factor Model*, the Five-Factor Model of Personality: Theoretical Perspectives, pp. 51-87, The Guilford Press, New York, NY, USA, 1996.

X. Wang, L. Zhang, Y. Lin, Y. Zhao, and X. Hu, "Computational models and optimal control

strategies for emotion contagion in the human population in emergencies,” *Knowledge-Based Systems*, vol. 109, pp. 35-47, 2016.

X. Guo, J. Chen, Y. Zheng, and J. Wei, “A heterogeneous lattice gas model for simulating pedestrian evacuation,” *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 3, pp. 582-592, 2012.

X.Z. Zheng, D. Tian, M. Zhang, C. Hu, and L. Tong, “A stairs evacuation model considering the pedestrian merging flows,” *Discrete Dynamics in Nature and Society*, vol. 2019, Article ID 7615479, 2019.

Y. Shin, S. Kim, and I. Moon, “Simultaneous evacuation and entrance planning in complex building based on dynamic network flows,” *Applied Mathematical Modelling*, vol. 73, pp. 545-562, 2019.

Y. Jiang, B. Chen, X. Li, and Z. Ding, “Dynamic navigation field in the social force model for pedestrian evacuation,” *Applied Mathematical Modelling*, vol. 80, pp. 815-826, 2020.

https://en.wikipedia.org/wiki/Beslan_school_siege.

https://en.wikipedia.org/wiki/School_shooting.

REFERENCES:

Bandman O.L. Cellular automaton models of spatial dynamics / O.L. Bandman // *System Informatics: collection of scientific tr./ ed. by A.G.Marchuk. – Novosibirsk: Publishing House of SB RAS, 2006. –C. 59-113.*

A.S. Brandrud, Michael Bretthauer, Guttorm Bratteb, et al., Local emergency medical response after a terrorist attack in Norway: a qualitative study, *Bmj Qual. Saf.* 26 (10) (2017).

B. Liu, Z. Liu, D. Sun, and C. Bi, “An evacuation route model of crowd based on emotion and geodesic,” *Mathematical Problems in Engineering*, vol. 2018, Article ID 6585102, 2018.

Burstedde, C. Simulation of pedestrian dynamics using a two-dimensional cellular automaton / C. Burstedde, K. Klauack, A. Schadschneider, J. Zittartz // *Physica A. – 2001. – Vol. 295. – P. 507-525.*

D. Helbing, I. Farkas, and T. Vicsek, “Simulating dynamical features of escape panic,” *Nature*, vol. 407, no. 6803, pp. 487-490, 2000.

H. Kim and S. Han, “Crowd evacuation simulation using active route choice model based on human characteristics,” *Simulation Modelling Practice and Theory*, vol. 87, pp. 369-378, 2018.

Kirchner, A. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics / A. Kirchner, A. Schadschneider // *Physica A. – 2002. – Vol. 312. – P. 260-276.*

M. Franovetter, “Threshold models of collective behavior,” *American Journal of Sociology*, vol. 83, no. 6, pp. 1420-1443, 1978.

M Hirsch, Pierre Carli, Remy Nizard, et al., The medical response to multisite ter-rorist attacks in Paris, *Lancet* 386 (10012) (2015) 2535-2538.

N. Ding, T. Chen, H Zhang, Experimental study of elevator loading and unloading time during evacuation in high-rise buildings, *Fire Technol.* 53 (1) (2017) 29-42.

N. Ding, H. Zhang, T. Chen, Simulation-based optimization of emergency evacuation strategy in ultra-high-rise buildings, *Natural Hazards* 89 (3) (2017) 1037-1184.

Nishinari, K. Extended floor field CA model for evacuation dynamics / K. Nishinari, A. Kirchner, A. Namazi, A. Schadschneider // *IEICE Transactions on Information and Systems*, E87-D. – 2004. – P. 726-732.

R.R. McCrae, P.T. Costa and J.S. Wiggins, *Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-Factor Model*, the Five-Factor Model of Personality: Theoretical Perspectives, pp. 51-87, The Guilford Press, New York, NY, USA, 1996.

X. Wang, L. Zhang, Y. Lin, Y. Zhao, and X. Hu, “Computational models and optimal control strategies for emotion contagion in the human population in emergencies,” *Knowledge-Based Systems*, vol. 109, pp. 35-47, 2016.

X. Guo, J. Chen, Y. Zheng, and J. Wei, “A heterogeneous lattice gas model for simulating

pedestrian evacuation,” *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 3, pp. 582-592, 2012.

X.Z. Zheng, D. Tian, M. Zhang, C. Hu, and L. Tong, “A stairs evacuation model considering the pedestrian merging flows,” *Discrete Dynamics in Nature and Society*, vol. 2019, Article ID 7615479, 2019.

Y. Shin, S. Kim, and I. Moon, “Simultaneous evacuation and entrance planning in complex building based on dynamic network flows,” *Applied Mathematical Modelling*, vol. 73, pp. 545-562, 2019.

Y. Jiang, B. Chen, X. Li, and Z. Ding, “Dynamic navigation field in the social force model for pedestrian evacuation,” *Applied Mathematical Modelling*, vol. 80, pp. 815-826, 2020.

https://en.wikipedia.org/wiki/Beslan_school_siege.

https://en.wikipedia.org/wiki/School_shooting.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 146-158
<https://doi.org/10.32014/2022.2518-1726.163>

ӨӘЖ 004.032.26:159.946.3(045)

А. Шаушенова^{1*}, А. Нурпейсова¹, Д. Досалянов², Г. Мауина¹

¹С. Сейфуллин атындағы Қазақ агротехникалық университеті,
Астана, Қазақстан;

²Нархоз университеті, Алматы, Қазақстан.
E-mail: Shaushenovsa_78@mail.ru

ПРОКТОРИНГ ЖҮЙЕСІНДЕ ЖАСАНДЫ НЕЙРОНДЫҚ ЖЕЛІЛЕРГЕ НЕГІЗДЕЛГЕН СӨЙЛЕУДІ ТАНУ МӘСЕЛЕЛЕРІ

Аннотация. Сөйлеуді тану міндеті – қазіргі заманның ең өзекті мәселелерінің бірі. Қазіргі уақытта көптеген дайын тану жүйелері бар болса да әртүрлі технологияларға негізделген сөйлеу, сөйлеуді тану міндеті толық шешілмеген, өйткені қолданыстағы жүйелерде белгілі бір кемшіліктер кездеседі. Атап айтқанда, жүйе жұмысының деректерді алмасу құралдарына қол жеткізуге тәуелділігі және тану дәлдігінің жеткіліксіздігі.

Сөйлеуді тану мәселелерін шешудегі перспективалық бағыттардың бірі жасанды нейрондық желілерді қолдану. Нейрондық желілер жалпылау қабілетіне байланысты үлгіні тану мәселелерінің әртүрлі кластарын шешуде кеңінен қолданылады.

Мақалада прокторинг автоматтандырылған жүйесінде сөйлеуді танудың жоғары сапалы стандартты модельдерінің бірі жасырын Марков модельдері (СММ) және нейрожелілік технологиялар сипатталады. Сөйлеуді танудың стандартты жүйесінің схемасы ұсынылып, әрбір кезеңдері сипатталып көрсетілді. Марков моделінің әлсіз тұстары аталып, сөйлеуді тануға арналған нейрондық желілердің негізгі түрлері қарастырылады. Сөйлеуді тану жүйелерінде қолданылатын нейрондық желілерге аналитикалық шолу жасалынған. Нейрондық желінің оңтайлы құрылымын таңдау және сөйлеуді тану жылдамдығын арттыру аспектілері анықталады.

Түйін сөздер: сөйлеуді тану, Марков модельдері, нейрондық желілер, нейрожелілік технологиялар.

А.Г. Шаушенова^{1*}, А.А. Нурпейсова¹, Д.Б. Досалянов², Г.М. Мауина¹

¹Казахский агротехнический университет имени С.Сейфуллина,
Астана, Казахстан;

²Университет Нархоз, Алматы, Казахстан.

E-mail: *Shaushenovsa_78@mail.ru*

ПРОБЛЕМЫ РАСПОЗНАВАНИЯ РЕЧИ НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В СИСТЕМЕ ПРОКТОРИНГА

Аннотация. Задача распознавания речи – одна из самых актуальных проблем нашего времени. Хотя в настоящее время существует множество готовых систем распознавания речи, основанных на различных технологиях, задача распознавания речи решена не полностью, поскольку существующие системы обладают определенными недостатками. В частности, зависимость работы системы от доступа к средствам обмена данными и недостаточная точность распознавания.

Одним из перспективных направлений в решении задач распознавания речи является использование искусственных нейронных сетей. Нейронные сети широко используются при решении различных классов задач распознавания образов благодаря их способности к обобщению.

В статье описываются скрытые Марковские модели (СММ) и нейросетевые технологии, одна из высококачественных стандартных моделей распознавания речи в автоматизированной системе прокторинга. Предложена схема стандартной системы распознавания речи с описанием каждого из этапов. Отмечаются слабые места марковской модели и рассматриваются основные типы нейронных сетей для распознавания речи. Проведен аналитический обзор нейронных сетей, используемых в системах распознавания речи. Определяются аспекты выбора оптимальной структуры нейронной сети и повышения скорости распознавания речи.

Ключевые слова: распознавание речи, Марковские модели, нейронные сети, нейросетевые технологии.

A.G. Shaushenova^{1*}, A.A. Nurpeisova¹, D.B. Dosalyanov², G.M. Mauina¹

¹Kazakh agrotechnical university named after S. Seifullin, Astana, Kazakhstan;

²Narxoz University, Almaty, Kazakhstan.

E-mail: *Shaushenovsa_78@mail.ru*

PROBLEMS OF SPEECH RECOGNITION BASED ON ARTIFICIAL NEURAL NETWORKS IN THE PROCTORING SYSTEM

Abstract. The task of speech recognition is one of the most pressing problems of our time. Although there are currently many ready-made speech recognition systems based on various technologies, the task of speech recognition is not fully solved, since existing systems have certain disadvantages. In particular, the dependence of the operation of the system on access to data exchange tools and insufficient recognition accuracy.

One of the promising directions in solving speech recognition problems is the use of artificial neural networks. Neural networks are widely used in solving various classes of pattern recognition problems due to their ability to generalize. The article describes hidden Markov models (CMM) and neural network technologies, one of the high-quality standard speech recognition models in an automated proctoring system. A scheme of a standard speech recognition system with a description of each of the stages is proposed. The weak points of the Markov model are noted and the main types of neural networks for speech recognition are considered. An analytical review of neural networks used in speech recognition systems is carried out. The aspects of choosing the optimal neural network structure and increasing the speed of speech recognition are determined.

Key words: speech recognition, Markov models, neural networks, neural network technologies.

Кіріспе. Covid-19 індеті және оған байланысты шектеулер мектептер мен жоғары оқу орындары үшін оқыту сапасын сақтауға және студенттердің білімін бағалауға бағытталған шұғыл шараларды қажет ететін үлкен проблемалар туғызды. Ақпараттық технологиялардың дамуымен онлайн оқыту және онлайн емтихандар кеңінен таралуда. Қашықтықтан оқыту және қашықтықтан емтихандар қабылдау белгілі бір дәрежеде оқытушылар мен студенттердің жұмысын жеңілдетті, сонымен бірге аймақтар арасындағы білім беру ресурстарындағы алшақтықты теңестірді. Алайда, қашықтықтан емтихан қабылдаудың белгілі бір кемшіліктері бар: бақылаушылардың болмауы қашықтықтан емтихан кезінде білім алушылардың көмекші материалдарды қолдануына мүмкіндік береді, бұл емтиханның әділдігі мен білім сапасына әсер етеді (Атоум т.б, 2017).

Жоғары мектепте ақпараттық технологияларды пайдалану және қашық-

тықтан білім беруді енгізу білім алушыларды бақылаудың нәтижелі іс-қимылдарын енгізуді талап етеді. Прокторинг бақылау, аралық және қорытынды жұмыстарды орындау кезінде тану, қадағалау және бағалау функцияларын орындайды. Бұл процесс адамның және ондағы машиналардың рөліне байланысты тірі, автоматтандырылған және жартылай автоматтандырылған болуы мүмкін. Жоғары мектепте Прокторинг бейімделу, стандарттау, ақпараттық қауіпсіздік, дербестендіру, интерактивтілік тезистеріне негізделуі тиіс (Daffin т.б, 2018). Бірқатар оқу орындары алаяқтықтың алдын алу үшін прокторинг технологиясы қажет деп санайды. Бірқатар басқа оқу орындары мен студенттер бұл тәсілге байланысты туындайтын қиындықтарға алаңдайды. Автоматтандырылған прокторинг бағдарламалары емтихан алушыларға алаяқтықтың алдын алу құралдарын ұсынады. Бұл бағдарламалар жүйелік ақпаратты жинай алады, желіге кіруді бұғаттайды және пернелерді басуды талдай алады. Олар сонымен қатар студенттер мен олардың айналасын жазу үшін компьютерлік камералар мен микрофондарды қолдана алады. Онлайн емтиханды бақылау процесінде сыртқы шу мен сөйлеуді айқындау да маңызды рөл атқарады (Fouad т.б, 2020).

Шу мен сөйлеуді тану жасанды интеллект санатына жатады және адамның физиологиялық сипаттамаларын идентификациялайтын модельдерді құру үшін микрофон, акустика, физикалық сенсорлар, биологиялық статистикалық принциптер және жетілдірілген математикалық әдістерді қолданады. Шу мен сөйлеуді тануды смартфондардағы қосымшалардан бастап «Ақылды үй» жүйелеріне дейін қарқынды дамыды. Сөйлеуді тану көптеген қосымшалардың интерфейсін басқару (навигаторлар, мессенджерлер және т.б.), телефон сөйлесулерін тану, сөйлеуді құру және т.б. салаларда қолданысқа ие (Moubtahij т.б, 2016). Сөйлеуді тану жүйесі бойынша келесі авторлардың еңбектерін ерекше атап көрсетуге болады: Б.М. Лобанов, Т.К. Винцюк, А.В. Фролов, Л.Р. Рабинер, Р.В. Шафер, У.А. Ли, Д.Х. Клетт, Хuedong D. Huang, Hsiao-Wuen Hon, Alex Astero. Жасанды интеллект технологиясының қарқынды дамуы және шу мен сөйлеуді тану қоғамдық жұмыста және өмірде кеңінен қолданыла бастады. Қазіргі уақытта компьютерлердің есептеу қуатының жоғарылауымен сөйлеуді тану барған сайын сұранысқа ие бола бастады (Алимурадов т.б, 2016).

Материалдар мен әдістер. Өкінішке қарай, сөйлеуді тану, диагностика және шуды басудың барлық мәселелерін шешуге жарамды әмбебап әдіс жоқ. Сондықтан сөйлеуді тану мәселелерін шешуге арналған құралдардың бай арсеналына және көптеген сәтті шешілген практикалық мәселелерге қарамастан, бұл тақырыпқа деген қызығушылық әлсіремейді. Бұл жаңа өндірістердің алуан түрлілігімен, нақты міндеттердің күрделілігімен, осы нақты міндеттерді дұрыс шешетін нейронды желілерді құру қажеттілігімен түсіндіріледі. мақалада зерттеу әдістемесі мәселені шешудің жаңалығымен және тиісті нәтижелермен анықталады.

Тапсырмаларды шешуде мәліметтер базасының теориясы, сигналдарды сандық өңдеу теориясы, деректерді өңдеу әдістері, сөйлеу мен шуды тану

бойынша ғылыми әдебиеттерді талдау; бақылау; мамандар қызметінің өнімдерін талдау нейрондық желілерді жобалау мен дамытудың заманауи жүйелері қолданылады.

Ғылымның осы бағыттарын қолдану алға қойылған мақсаттарға жету жолдарын және таңдалған тәсілдерді негіздеуді дәлелдейді.

Талқылау. Сөйлеу сигналдарын өңдеу – бұл ақпаратты сүзу, күшейту және алу, кодтау, қысу және қалпына келтіру жүзеге асырылатын ғылым саласы. Сөйлеуді тану жүйелерінде өңдеу келесі міндеттерді қамтиды:

- сүзу және шуды басу;
- ақпараттық бөлімдерге сегментация;
- ақпараттық параметрлерді анықтау;
- тану.

Сөйлеуді танудың жоғары сапалы стандартты модельдері бар, олар жақсы нәтиже көрсетеді және көптеген түрлі бөліктерден тұрады. Бірақ оларда барлық компоненттер өздігінен оқытылады, ал кейбір компоненттердегі қателер басқаларында қателіктер тудыруы мүмкін. Стандартты жүйенің сценарийі көптеген қадамдардан тұрады, олар гигабайттарды сақтауды қажет етеді.

Сонымен қатар, акустикалық модельдеу үшін кеңінен қолданылатын жасырын Марков модельдерін (СММ) қолданудың кемшіліктері бар: бақылаулар тәуелсіз кездейсоқ шамалар болып табылады және "әлсіз модельдер" қолданылады — бірінші ретгі Марков модельдері (Amrānīa т.б, 2016). Сөйлеуді танудың стандартты жүйесі. Сөйлеуді автоматты түрде танудың мақсаты S дыбыстық сигналын W сөз тізбегіне түрлендіру. Бұл тапсырманы S кіріс сигналы арқылы сөздердің ең ықтимал тізбегін іздеу ретінде тұжырымдауға болады:

$$W^* = \arg \max_{W} P(W | S),$$

мұндағы W — гипотезалар жиыны.

Әдетте, сөйлеуді тану жүйесі 1-суретте көрсетілгендей тапсырманы үш кезеңге бөледі: белгілерді таңдау, акустикалық модельдеу және тізбекті декодтау. Әр қадамды толығырақ қарастырайық.



Сурет 1. Сөйлеуді танудың стандартты жүйесі

Белгілерді таңдау. Бұл қадамда тапсырмаға байланысты S сөйлеу сигналынан X белгілерін таңдау, сөйлеу ерекшеліктері және белгілер векторларының кеңістігін азайту орындалады. Белгілердің ең танымал екі түрі: аз жиілікті кепстральды коэффициенттер (MFCC) және сызықтық болжаудың перцептивті коэффициенттері (perceptual linear prediction cepstral coefficient; PLP). Бұл белгілерді алу келесі кезеңдерден тұрады:

1. дауыс сигналының сегментін көптеген жиіліктерге түрлендіру (мысалы, Дискретті Фурье түрлендіруі арқылы);
2. түрлі сүзгілерді қолдану;
3. сызықтық емес функцияны қолдану ($\ln(\cdot)$, $\sqrt{\cdot}$);
4. корреляцияланбаған белгілердің өлшемін азайту үшін әртүрлі түрлендірулерді қолдану (дискретті косинустық түрлендіру немесе авторегрессиялық модельдер).

Акустикалық модельдеу. Акустикалық модельдеу белгілер мен лингвистикалық бірліктер арасында статистикалық тәуелділікті құру үшін қолданылады, мысалы фонемалармен.

Тізбекті декодтау. Тізбекті декодтау X белгілерінің тізбегін W сөз тізбегіне айналдырады. Бұл қадамды келесідей сипаттауға болады:

$P_L(W)$ - тіл модельдері арқылы алынған априорлық ықтималдық;

$PA(X|W)$ - акустикалық модельдерге негізделген ықтималдылық функциясы (Гусев т.б., 2013).

Нәтижелер. Кейінгі кездері сөйлеуді тану жүйесінде нейрондық желілер кеңінен таралды. Әдетте терең жасанды нейрондық желілер (ins) интегралды модель ретінде қызмет етеді. Бұл тәсілдің стандарт модельден артықшылықтарын атап өтейік:

- интегралды модельдерді енгізу оңайырақ, өйткені олар тек бір фреймворкты қолдана отырып жазуға болатын бір нейрондық желіні қамтуы мүмкін, бұл бағдарлама кодындағы қателіктердің туындау ықтималдығын азайтады;

- интегралдық модельдер жақсы өнімділікті көрсетеді (жылдамдық, кейде дәлдік);

- бұл интегралды модельдер компьютердің жадының аз мөлшерін қажет етеді, бұл оларды жергілікті мобильді құрылғыларда пайдалануға мүмкіндік береді (Jha, 2008).

Мұндай модельдердің кемшілігі оқу үшін көптеген белгіленген мәліметтерді қажет етуінде, бұл кейбір тапсырмалар түрлерінде қиындық тудыруы мүмкін.

Адамның сөйлеуі әркезде өзгеріп отыруымен сипатталатыны белгілі. Бұл бірнеше себептерге байланысты. Біріншіден, тіпті бір сөйлеп тұрған адамның акустикалық бірліктерінің орындалуы олардың спектрлік құрамы мен айтылу ұзақтығы әр түрлі болады. Бұл адамның эмоционалды жағдайының өзгеруіне, оның сол уақыттағы жай-күйіне байланысты болуы мүмкін. Екіншіден, коартикуляциялық әсерлердің болуы сөздер мен фонемалардың айтылуы олардың контекстіне байланысты болатындығына әкеледі. Үшіншіден,

әртүрлі сипаттағы кедергілер сөйлеу сигналының өзгеруіне әкеледі. Осы факторлардың барлығын және басқа да бірқатар шектеулерді ескеріп, нақты уақыттағы сөйлеуді жоғары сапалы тану үшін жоғары жылдамдықты есептеу құралдары қажет екенін атап өткен жөн. Бұл талаптарды төмендетудің бір тәсілі нейрокомпьютерлерде жасанды нейрон желілілерін іске асыруда параллель есептеулерді қолдану (Maniag т.б., 2021).

Сөйлеудің жеке айтылған сөздерді танудың қарапайым схемасын қарастырамыз (2-сурет).



Сурет 2. Жеке айтылған сөздерді тану схемасы

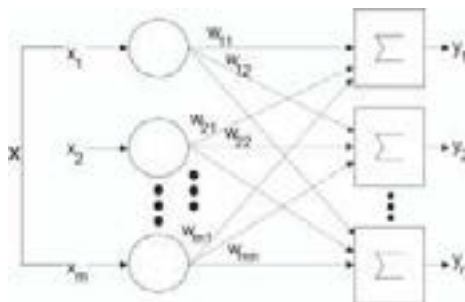
Бұл жағдайда тану процесін үш кезеңге бөлуге болады. Бірінші кезеңде акустикалық препроцессор кіріс сөйлеу сигналын белгіленген уақыт аралығында алынған белгілер векторларының немесе акустикалық векторлардың тізбегіне түрлендіреді. Әдетте, бұл векторларда сөйлеу сигналының қысқа сегменттерін сипаттайтын спектрлік немесе кепстральды коэффициенттер болады.

Екінші кезеңде векторлар сөз модельдеріндегі сілтемелермен салыстырылады және олардың жергілікті өлшемдері немесе сәйкестік өлшемдері есептеледі (жалпы жағдайда бірнеше сипаттамалық векторлармен ұсынылған сөйлеу сегменттері салыстырылады). Үшінші кезеңде бұл өлшемдер белгілер векторларының тізбегін сөз модельдерін құрайтын эталондар тізбегімен уақытша туралау үшін қолданылады және сөздердің сәйкестік шаралары есептеледі. Уақытша туралау - сөйлеу жылдамдығындағы өзгерістерді өтеу үшін қолданылады. Барлық осы әрекеттерді орындағаннан кейін, танушы сәйкестік өлшемі максималды болатын сөзді тандайды. Біріктірілген сөйлеуді тану кезінде есептеудің екінші кезеңінде алынған жергілікті өлшемдер уақытша туралау және жеке сөйлемдер мен мәлімдемелер үшін сәйкестік шараларын анықтау үшін қолданылады (Sundermeyer, 2011).

2-суретте көрсетілген тану схемасында Нейрондық желілер (НЖ) жергілікті өлшемдерді есептеу кезінде есептеудің екінші кезеңінде сәтті қолданылады.

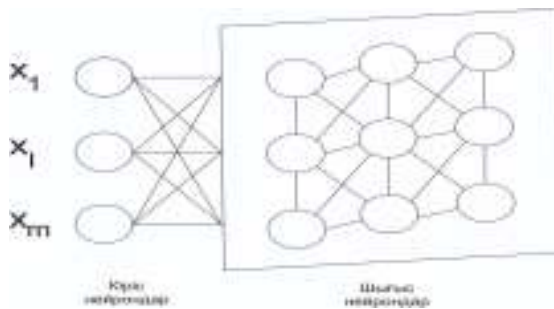
Үздіксіз бақыланатын статистикалық танушылар үшін бұл өлшемдер белгілер векторларының ықтималдылық функциясының монотонды функциялары болып табылады.

Дискретті бақыланатын сөйлеуді танушылар алдымен векторлық кванттауды орындайды және белгілердің әр векторына код кітабынан белгілі бір таңбаны тағайындайды. Содан кейін, осы таңбалардың негізінде әр сілтеме векторы үшін таңбаларды бақылау ықтималдығы бар арнайы кестелерді қолдана отырып, жергілікті өлшемдер есептеледі. Мұндай есептеулерді сызықтық түйіндерден тұратын бір қабатты перцептрондармен жасауға болады (3-сурет), олардың саны стандарттар санына тең. Мұндай перцептронның кіріс саны мүмкін таңбалар санына тең болуы керек.



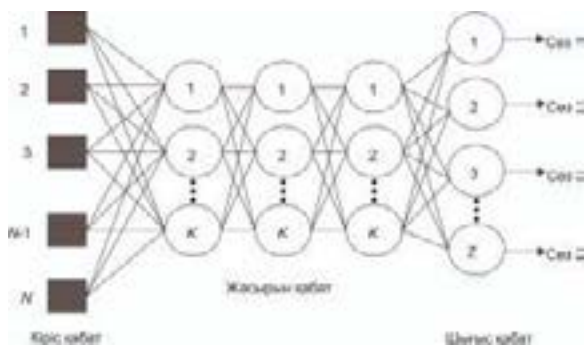
Сурет 3. Бір қабатты перцептрон

Векторлық кванттауды Кохоненнің белгілер картасы сияқты желі арқылы жасауға болады (4-сурет). Мұндай желі -бұл әр мүмкін символға бір түйіннен тұратын код кітабының түйіндерінің екі өлшемді массиві. Әрбір түйін желінің кіріс векторы мен түйіннің салмақтарымен ұсынылған сәйкес сілтеме арасындағы Евклид қашықтығын есептейді, содан кейін ең аз Евклид қашықтығы бар түйін таңдалады. Бұл желінің салмағы кохонен алгоритмін, оның модификацияларын немесе Евклид қашықтығын метрика ретінде қолданатын кез-келген басқа дәстүрлі векторлық кванттау алгоритмін қолдана отырып есептеледі (мысалы, k-орташа алгоритмді қолдана отырып).



Сурет 4. Кохоненнің нейрондық желісі

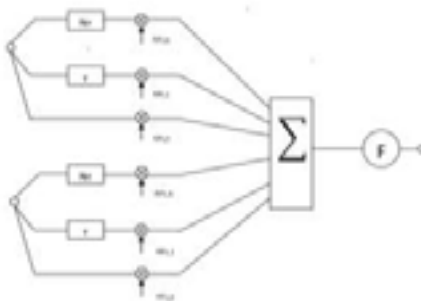
Нейрондық желі көп қабатты нейрондық желілерді (5-сурет) танудың бастапқы кезеңінде препроцессор шығаратын белгілер векторларының мөлшерін азайту үшін де қолдануға болады. Мұндай НЖ кірістермен және жасырын түйіндердің бір немесе одан да көп қабаттарымен бірдей шығуларға ие. НЖ жаттығу кезінде оның салмағы жасырын түйіндердің кішкене қабаты арқылы кез-келген кіріс векторын шығару кезінде көбейте алатындай етіп таңдалады. Желіні оқытқаннан кейін бұл түйіндердің шығуы сөйлеуді одан әрі өңдеу және тану үшін аз өлшемді кіріс векторлары ретінде пайдаланылуы мүмкін (Mitra т.б., 2016). НЖ фонемалардың статикалық кескіндерін, буындарды және окшауланған сөздердің шағын сөздіктерін жіктеу үшін қолданылған жағдайда, оны жүзеге асырудың стационарлық бөлігін сипаттайтын белгілер векторы кіріс кескіні ретінде таңдалуы мүмкін.



5 сурет. Сөйлеуді тануға арналған көп қабатты нейрондық желі құрылымы

Сөйлеуді тану үшін арнайы жасалған және қысқа уақыттық кідірістер мен уақытша интеграцияны немесе қайталанатын байланыстарды жүзеге асыратын түйіндерді қамтитын динамикалық нейрондық желі жіктеуіштері ерекше қызығушылық тудырады. Әдетте, мұндай жіктеуіштер оқыту мен бақылау үлгілерінің аз уақыттық ауысуына сезімтал емес, сондықтан жоғары сапалы жұмыс үшін сөйлеу деректерін дәл сегментациялауды қажет етпейді. Сөйлеуді тану кезінде динамикалық желілерді қолдану статикалық желілерге тән негізгі кемшіліктерді жеңуге мүмкіндік береді және эксперименттік зерттеулер көрсеткендей, акустикалық ұқсас сөздер, дауыссыз және дауысты дыбыстар үшін танудың жоғары сапасына әкеледі (Newton, 2018).

Уақытша кешіктіруі бар Нейрон желісі (УКНЖ) тораптары уақытша кідіруге өзгерту енгізген көпқабатты перцептрон білдіреді. 6-суретте N кідірісі бар $\tau, 2\tau, \dots, N\tau$ торап көрсетілген.

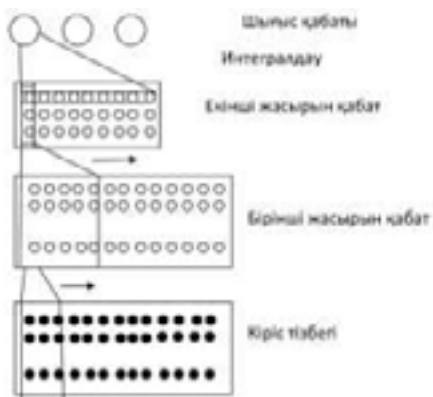


Сурет 6. Кідірістері бар нейрондық желі түйінінің схемасы

Ол $N+1$ -де алынған J кірістерінің дәйекті моменттерін тиісті салмақ коэффициенттеріне көбейтіп, шекті алып тастайды және алынған нәтиженің сызықты емес F функциясын есептейді.

Үш фонеманы (немесе фонеманың үш класын) тану үшін ұсынылған үш қабатты УКНЖ архитектурасы 6 суретте көрсетілген (ол тек бір шығыс түйініне арналған байланыстарды көрсетеді) (Varble т.б., 2014).

6-суретте акустикалық векторлардың кіріс тізбегін өңдеу төменгі деңгейдегі түйіндердің суреттерінен уақытша кідірістер терезелерінің өтуіне тең екендігі көрсетілген. Ең төменгі деңгейде бұл кескіндер сенсорлық кірістен, яғни акустикалық векторлардан тұрады, желінің жасырын қабаттарының түйіндері қозғалмалы белгілер детекторлары болып табылады және кіріс тізбегінің кез-келген жерінде қажетті суреттерді анықтай алады. Шығу түйіндерінің екінші қабатпен байланысы бірдей болғандықтан, мұндай детекторлар үшін кез-келген уақыт нүктелері тең. Бұл желіні фонемалардың оқыту және бақылау үлгілерінің уақыттық ауысуларына инвариантты етеді (бұл ауысулар соншалықты үлкен болмаған жағдайда, маңызды негізгі белгілер желінің кіріс тізбегінен тыс болады) (Карпов, 2012).



Сурет 7. Уақытша кідірістері бар нейрондық желі архитектурасы

Сөйлеуді тану кезінде басқа функцияларды орындауға арналған нейрондық желілік құрылымдарға шолуды басқа әдебиеттерден табуға болады.

Қортынды. Бұл мақалада сөйлеуді тану үшін жоғары сапалы стандартты модельдер - жасырын Марков модельдері (СММ) және нейрондық желілердің әртүрлі құрылымдарын пайдалану мүмкіндіктері қарастырылды. СММ-де барлық компоненттер өздігінен оқытылғанымен, кейбір компоненттердегі қателер басқаларында қателіктер тудыруы мүмкін және жадыда үлкен көлемді орындарды талап етеді. Нейронды желілердің қарқынды дамуы сөйлеуді тану жүйесі саласында көптеген жетістіктерге жетуге жол ашты. Соның ішінде Нейрондық желі әдістері параллелизация есептеулерінің арқасында тану жылдамдығын арттырады деп қорытынды жасауға болады. Көп қабатты нейрондық желілерді қолдану нейрондар арасындағы синаптикалық байланыстардың санын, есте сақтау шығындарын, оқуға кететін уақытты және танымал сөздердің сөздігін кеңейту кезінде желінің жұмысын көбейту мәселесін шешеді. Осының арқасында сөйлеуді тану жүйесі нақты уақыт режимінде жұмыс жасай алады.

Бұл ғылыми мақала ИРН №АР09259657 «Студенттердің білімін бақылауға арналған прокторинг автоматтандырылған жүйесін зерттеу және әзірлеу» жобасы бойынша дайындалды.

Information about the authors:

Shaushenova Anargul – Candidate of Technical Sciences, Head of the Department of Information Systems, S. Seifullin Kazakh Agrotechnical University, Astana, Kazakhstan, Shaushenova_78@mail.ru, <https://orcid.org/0000-0002-3164-3688>;

Nurpeisova Ardak – Master of mathematics, Senior teacher of the department of information and communication technologies, S. Seifullin Kazakh Agrotechnical University, Astana, Kazakhstan, nurpeisova.ardak81@gmail.com, <https://orcid.org/0000-0002-1245-8313>;

Dossalyanov Damir – doctor of Philosophy PhD, Public and local management, Narxoz University, Almaty, Kazakhstan, ms_018@mail.ru, <https://orcid.org/0000-0001-9796-4049>;

Mauina Gulalem – Master of Engineering and Technology, Assistant of the Department of “Information Systems”, S. Seifullin Kazakh Agrotechnical University, Astana, Kazakhstan, alema85@mail.ru, <https://orcid.org/0000-0000-1975-36781>.

ӘДЕБИЕТТЕР:

Atoum Y., Chen L., Liu A.X., Hsu S.D.H., Liu X. 2017. Automated Online Exam Proctoring. IEEE Transactions on Multimedia. 19(7), 7828141, P. 1609-1624

Daffin L.W., Jones A.A. (2018). Comparing student performance on the proctored and nonproctored exams in online psychology courses. Online Learning, 22 (1), P.131–145.

El Amrania M.Y. Building CMU Sphinx language model for the Holy Quran using simplified Arabic phonemes / M.Y. El Amrania, M.M. Hafizur Rahmanb, M.R. Wahiddinb, A. Shahb // *Egyptian Informatics Journal* – 2016. – V. 17. №3. – P. 305–314.

El Moubtahij H. Using features of local densities, statistics and HMM toolkit (HTK) for offline Arabic handwriting text recognition / H. El Moubtahij, A. Halli, K. Satori // *Journal of Electrical Systems and Information Technology* – 2016. – V. 3. №3. – P. 99-110.

Fouad Y., Lodder A., Hurdey J., et al. (2018) A lawful basis for online proctoring. Available at: https://proctorexam.com/wp-content/uploads/2020/02/OP4RE_A0-Lawful-basis-for-Online-Proctoring-vFinal-1.pdf.

Jha M. Improved unsupervised speech recognition system using MLLR speaker adaptation and confidence measurement / M. Jha et al. // *V Jornadas en Tecnologias del Habla (VJTH'2008)* – 2008. – P. 255-258.

Maniar S., Sukhani K., Shah K., Dhage S. 2021. Automated Proctoring System using Computer Vision Techniques. 2021 International Conference on System, Computation, Automation and Networking, ICSCAN 2021.

Mitra S., Gofman M.I. 2016. Towards Greater Integrity in Online Exams. Assoc Informat Syst. AMCIS 2016 PROCEEDINGS

Newton, P. (2018). How common is commercial contract cheating in higher education? *Front. Educ.* 3, 1–18. doi: 10.3389/fevue.2018.00067.

Sundermeyer The rwth 2010 quaero asr evaluation system for english, french, and german / M. Sundermeyer et al. // *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)* – 2011. – P. 2212-2215.

Varble D.L., Haute T. (2014). Reducing cheating opportunities in online tests. *Atlantic Marketing Journal*, 3 (3), P.131–149.

Алимурадов А.К. Адаптивный метод повышения эффективности голосового управления / А.К. Алимурадов, П.П. Чураков // *Труды Международной научно-технической конференции «Перспективные информационные технологии»* – 2016. – С. 196-200.

Гусев М.Н. Система распознавания речи: основные модели и алгоритмы / М.Н. Гусев, В.М. Дегтярев. – СПб.: Знак, 2013. – 128 с.

Карпов А.А. Методология оценивания работы систем автоматического распознавания речи / А.А. Карпов, И.С. Кипяtkова // *Известия высших учебных заведений. Приборостроение.* – 2012. – Т. 55. – №. 11. – С. 38-43.

REFERENCES:

Atoum Y., Chen L., Liu A.X., Hsu S.D.H., Liu X. 2017. Automated Online Exam Proctoring. *IEEE Transactions on Multimedia.* 19(7), 7828141, P. 1609-1624

Daffin L.W., Jones A.A. (2018). Comparing student performance on the proctored and nonproctored exams in online psychology courses. *Online Learning*, 22 (1), P.131–145.

El Moubtahij H. Using features of local densities, statistics and HMM toolkit (HTK) for offline Arabic handwriting text recognition / H. El Moubtahij, A. Halli, K. Satori // *Journal of Electrical Systems and Information Technology* – 2016. – V. 3. №3. – P. 99-110.

El Amrania M.Y. Building CMU Sphinx language model for the Holy Quran using simplified Arabic phonemes / M.Y. El Amrania, M.M. Hafizur Rahmanb, M.R. Wahiddinb, A. Shahb // *Egyptian Informatics Journal* – 2016. – V. 17. №3. – P. 305–314.

Fouad Y, Lodder A, Hurdey J, et al. (2018) A lawful basis for online proctoring. Available at: https://proctorexam.com/wp-content/uploads/2020/02/OP4RE_A0-Lawful-basis-for-Online-Proctoring-vFinal-1.pdf

Jha M. Improved unsupervised speech recognition system using MLLR speaker adaptation and confidence measurement / M. Jha et al. // *V Jornadas en Tecnologias del Habla (VJTH'2008)* – 2008. – P. 255-258.

Maniar S., Sukhani K., Shah K., Dhage S. 2021. Automated Proctoring System using Computer

Vision Techniques. 2021 International Conference on System, Computation, Automation and Networking, ICSCAN 2021.

Mitra S., Gofman M.I. 2016. Towards Greater Integrity in Online Exams. Assoc Informat Syst. AMCIS 2016 PROCEEDINGS.

Newton P. (2018). How common is commercial contract cheating in higher education? Front. Educ. 3, 1–18. doi: 10.3389/fedue.2018.00067.

Sundermeyer The rwth 2010 quaero asr evaluation system for English, French, and German / M. Sundermeyer et al. // Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP) – 2011. – P. 2212-2215.

Varble D.L., Haute T. (2014). Reducing cheating opportunities in online tests. Atlantic Marketing Journal, 3 (3), P.131–149.

Alimuradov A.K. Methodus adaptiva augendi vocis potestatem augendi/A.K. Alimuradov, P.P. Churakov//Acta Conferentiarum Scientificorum et technicorum Internationalium “Perspectivae Informationes Technologies” - 2016. - P. 196-200.

Gusev M.N. Systema recognitionis locutionis: exempla fundamentalia et algorithms / M.N. Gusev, V.M. Degtyarev. – Petropoli:Sign, 2013. – 128 p.

Karpov A.A. Methodus iudicandi opus systematis latis elocutionis recognitionis / A.A. Karpov, I.S. Kipyatkova // News altiorum institutionum educationis. Instrumentum. - 2012. - T. 55. - No. 11

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X

Volume 4, Number 344 (2022), 159-167
<https://doi.org/10.32014/2022.2518-1726.164>

УДК 004.931

А.Ә. Шекербек[†], Г.Б. Абдикеримова¹, Ә.М. Сабыр², Ж.С. Әбілқайыр²

¹Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

²Қазақ технология және бизнес университеті, Астана, Қазақстан.

E-mail: shekerbek80@mail.ru

КЕУДЕ КЛЕТКАСЫНЫҢ ПАТОЛОГИЯСЫН АНЫҚТАУ ҮШІН ӘДІС ПЕН АЛГОРИТМДІ ҚОЛДАНУ

Аннотация. Кескінді талдауда маңызды сипаттама ұшақтар мен спутниктік құрылғыларды пайдалану арқылы алынғандардан бастап биомедициналық зерттеулердегі микроскопиялықтарға дейін барлық кескіндерде болатын текстура болып табылады. Кеуде қуысының рентгенографиясы қатерлі ісік, пневмония және туберкулез сияқты өкпенің ауыр ауруларын зерттеу үшін ең жиі қолданылатын және тиімді диагностикалық әдіс. Рентгеннің ең үлкен объектісі – өкпе. Өкпе контурларының пішіні мен өлшемін дұрыс бөлу диагностика үшін маңызды ақпарат, оның негізінде интеллектуалды ақпараттық органы құруға болады. Рентгенографияны қолданудың орындылығына қарамастан, диагностика сатысында ауру анықталмауы мүмкін. Осыған байланысты патологияның даму қаупі артып, өлімге жағдайына әкелуі мүмкін.

Мақалада ең дәл нәтиже алу үшін автокорреляция функциясын қолдану арқылы пневмония ауруларын кластерлеу мәселесін шешу ұсынылады. Осылайша, өкпе рентгенографиясын диагностикалауда дәрігерге сенімді көмекші құрал қалыптасады. Дұрыс жұмыс істейтін автоматтандырылған жүйені құруда кескінді алдын ала өңдеу және дерекқорды құру маңызды рөл атқарады. Ең жиі кездесетін аурулар пневмония, бронхопневмония, полисегментальды пневмония және т.б. Сондықтан классификациялық тапсырма үшін екі кластың суреттері таңдалды: сау және пневмония түрлері. Бұл құжат текстуралық ерекшеліктерді және k-орталардың жіктелуін есептеу негізінде өкпенің рентгендік кескіндеріндегі қызығушылық аймақтарын бөлектеу үшін автокорреляция функциясының қолданылуын көрсетеді.

Түйін сөздер: кеуде қуысының рентгенографиясы, медициналық бейнелеу, текстура, кластерлеу, патология.

A. Shekerbek^{1*}, G. Abdikerimova¹, A. Sabyr², Zh. Abilkaiyr²

¹Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan;

²Kazakh University of Technology and Business, Astana, Kazakhstan.

E-mail: *shekerbek80@mail.ru*

APPLICATION OF THE METHOD AND ALGORITHM FOR THE DETECTION OF CHEST PATHOLOGY

Abstract. In image analysis, an important characteristic is a texture, which is present in all images, from those obtained using aircraft and satellite devices to microscopic ones in biomedical research. Chest X-ray is the most commonly used and effective diagnostic method for investigations of serious lung diseases such as cancer, pneumonia, and tuberculosis. The lungs are the largest object of the x-ray. The correct separation of the shape and size of the contours of the lungs is important information for the diagnosis, based on which an intelligent information environment can be created. Despite the feasibility of using radiography, at the stage of diagnosis, there is a possibility that the disease may not be detected. In this regard, there is a risk of developing pathology, which can lead to death.

The article proposes a solution to the problem of clustering pneumonia diseases by using an autocorrelation function to obtain the most accurate result. Thus, a reliable auxiliary tool will be formed for the doctor in the diagnosis of pulmonary radiographs. Image pre-processing and database generation play an important role in creating a correctly functioning automated system. The most common diseases are pneumonia, bronchopneumonia, polysegmental pneumonia, etc. Therefore, for the classification task, images of two classes were selected: healthy, and types of pneumonia. This paper shows the applicability of the autocorrelation function to highlight areas of interest in x-ray images of the lungs, based on the calculation of textural features and k-means classification.

Key words: chest radiograph, medical imaging, texture, clustering, pathology.

А.А. Шекербек^{1*}, Г.Б. Абдикеримова¹, А.М. Сабыр², Ж.С. Абулхаир²

¹Евразийский национальный университет имени Л.Н. Гумилева,
Астана, Казахстан;

²Казахский университет технологий и бизнеса, Астана, Казахстан.
E-mail: *shekerbek80@mail.ru*

ПРИМЕНЕНИЕ МЕТОДА И АЛГОРИТМА ДЛЯ ВЫЯВЛЕНИЯ ПАТОЛОГИИ ГРУДНОЙ КЛЕТКИ

Аннотация. При анализе изображений важной характеристикой является текстура, которая присутствует на всех изображениях: от полученных с

помощью авиационных и спутниковых приборов до микроскопических в биомедицинских исследованиях. Рентген грудной клетки является наиболее часто используемым и эффективным методом диагностики серьезных заболеваний легких, таких как рак, пневмония и туберкулез. Легкие – самый крупный объект рентгена. Правильное разделение формы и размеров контуров легких является важной информацией для диагностики, на основе которой может быть создана интеллектуальная информационная среда. Несмотря на целесообразность использования рентгенографии, на этапе диагностики существует вероятность того, что заболевание может быть не выявлено. В связи с этим существует риск развития патологии, которая может привести к летальному исходу.

В статье предлагается решение проблемы кластеризации пневмоний за счет использования автокорреляционной функции для получения наиболее точного результата. Таким образом, у врача будет сформирован надежный вспомогательный инструмент в диагностике рентгенограмм легких. Предварительная обработка изображений и формирование базы данных играют важную роль в создании корректно функционирующей автоматизированной системы. Наиболее распространенными заболеваниями являются пневмония, бронхопневмония, полисегментарная пневмония и др. Поэтому для задачи классификации были выбраны изображения двух классов: здоровых и видов пневмонии. В этой статье показана применимость автокорреляционной функции для выделения областей интереса на рентгеновских изображениях легких на основе расчета текстурных признаков и классификации k-средних.

Ключевые слова: рентгенограмма грудной клетки, медицинская визуализация, текстура, кластеризация, патология.

Кіріспе. Радиациялық диагностика әдістерінің таралуымен рентгенологтардың жұмыс көлемі айтарлықтай өсті. Денсаулық сақтаудағы цифрландырудың үздіксіз процесі медицинаның әртүрлі салаларында көбірек жаңа технологияларды қолдануға мүмкіндік береді. <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia> дерек-қорынан алынған деректер рентгенологтың жұмыс тәжірибесіне жасанды интеллектті енгізу мүмкіндігін одан әрі бағалау үшін патологияларды анықтау үшін математикалық әдістермен тексерілді. жұмыс көлемін және ресурстар шығындарын азайту үшін флюорограммалармен жұмысты оңтайландыру. Денсаулыққа қатысты цифрлық деректер анағұрлым айқын және дәстүрлі деректерден, мысалы, медициналық жазбалардағы жазбалардан, күнделікті өміріміз туралы кейде анық емес ақпаратқа, сондай-ақ біз өмір сүретін органы сипаттайтын деректердің кең ауқымына дейін кеңейеді. Денсаулық сақтаудағы цифрландырудың үздіксіз процесі медицинаның әртүрлі салаларында көбірек жаңа технологияларды қолдануға мүмкіндік береді (Pisaniello & Dixon, 2020). Радиациялық диагностика әдістерінің таралуымен рентгенологтардың жұмыс көлемі айтарлықтай өсті. 2019 жылы орташа рентгенолог сұранысты

қанағаттандыру үшін 8 сағаттық жұмыс күні ішінде әрбір 3-4 секунд сайын флюорограмманы, рентгенографияны немесе бір КТ немесе МРТ суретін түсіндіруі керек деп есептелді. Түсіндіруді қажет ететін суреттер санының артуы жұмыс көлемінің айтарлықтай өскенін білдіреді. Технологиялық жетістіктермен радиологтар бір емтиханда көбірек кескіндерді өңдейді. Жұмыс уақытын қысқарту рентгенологқа түсетін ауыртпалықты азайту ғана емес, сонымен қатар ресурстық шығындарды азайту, сол арқылы денсаулық сақтаудағы экономикалық жағдайды жақсарту үшін үлкен қызығушылық тудырады (Driver және т.б., 2020). Кеуде қуысының флюорографиясы әлемдегі ең жиі қолданылатын рентгендік әдістердің бірі болып табылады, қазіргі уақытта патологияны цифрландыру үшін зерттеушілер қол жетімді.

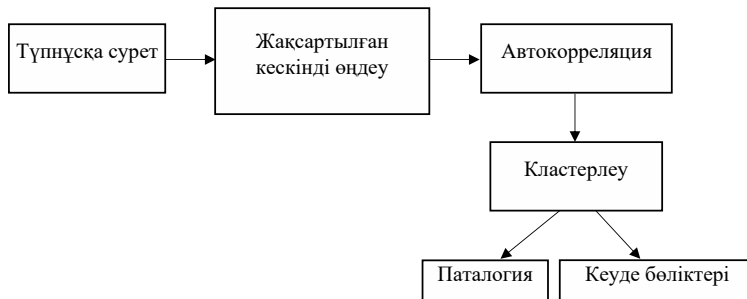
Зерттеу әдісі. Бұл тәсілде текстура кескінің туынды емес элементтерінің тонының кеңістіктік өлшемімен байланысты (тонның туынды емес элементі – белгілі бір «аномальды» белгілері бар кескін аймағы). Автокорреляция функциясының мәні біртекті туынды емес элементтердің өлшемін сипаттайтын белгі ғана. Кеңістіктік орналасу корреляция коэффициентімен сипатталады, ол бір сурет элементінің жарықтылығының басқасының жарықтығына сызықтық тәуелділігінің өлшемі болып табылады (Haralick, 1979). Математикалық тұрғыдан автокорреляциялық функцияны есептеу процесін келесі формуламен сипаттауға болады:

$$\rho(x, y) = \frac{\frac{1}{(L_x - |x|) \times (L_y - |y|)} \iint S(u, v) \times S(u+x, v+y) du dv}{\frac{1}{L_x \times L_y} \iint S^2(u, v) du dv},$$

мұндағы $S(u, v)$ - (u, v) координаталары бар нүктедегі мөлдірліктердің мөлдірлігі, (x, y) - x және y бағыттарындағы ығысу шамасы. Сондай-ақ $0 \leq u \leq L_x$, $0 \leq v \leq L_y$ тіктөртбұрышының сыртында мөлдірлік нөлге тең деп есептейміз.

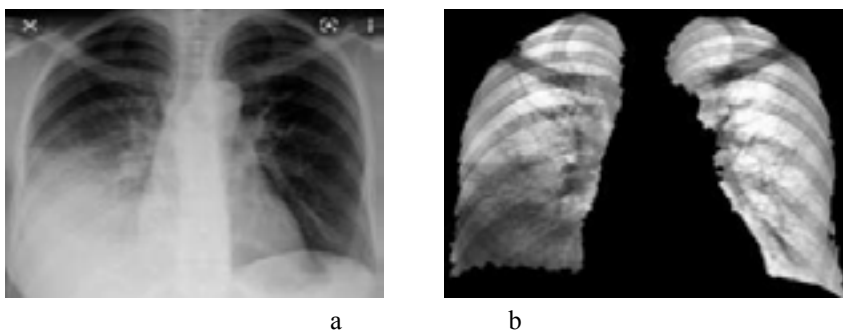
Нәтижелер мен талқылау. Зерттеу барысында келесі нәтижелер алынды: Python тілінде енгізілген бағдарлама 15 (100%) флюорограмманы «қалыпты», 17 (99%) флюорограмманы «патология» және 1 (1%) «дабыл» деп белгіледі. Автокорреляция функциясының өңдеу деректері мен рентгенологтардың қорытындыларының деректерін салыстыру кезінде бағдарлама мен дәрігерлердің нәтижелері толығымен сәйкес келетіні анықталды. Алынған нәтижелер қолданылатын әдістің жоғары диагностикалық дәлдігін, сондай-ақ радиологтың жұмысын автоматтандыру әдісін қолдану мүмкіндігін көрсетеді.

Есептеу барысында (<https://www.kaggle.com>) деректер қорынан алынған суреттер қарастырылды. Барлық кескіндер (1) суретке сәйкес өңделді.



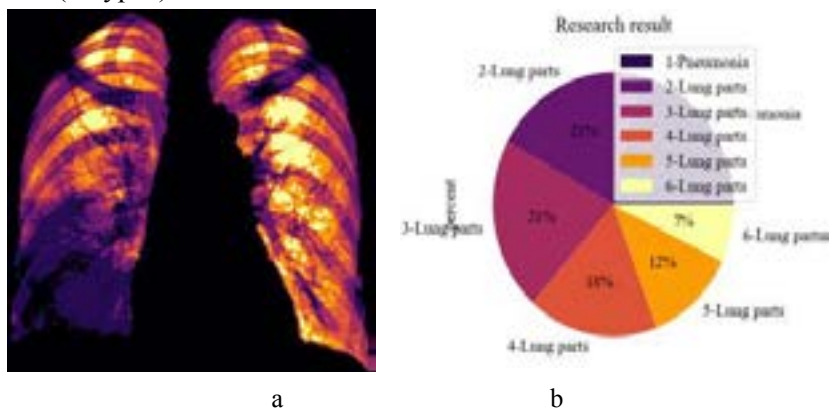
Сурет 1. Stages of determining the pathology of the chest

Бағдарлама Python тілінде жүзеге асырылды. 1-суреттегі алгоритм бойынша жүргізілген зерттеу жұмысы мәліметтер қоры кескіндерінің патологиясын анықтауға бағытталған. Жұмыс барысында түпнұсқа суреттер өңделді, яғни зерттеуге тек өкпе қуысының суреттері таңдалып, контраст күшейтілді (2-сурет).



2-сурет. Түпнұсқа кескін (a), түпнұсқа кескінді жақсарту (b)

Әрбір суретте «аномальды» текстуралар ажыратылып, олардың пайызы анықталды (3-сурет).



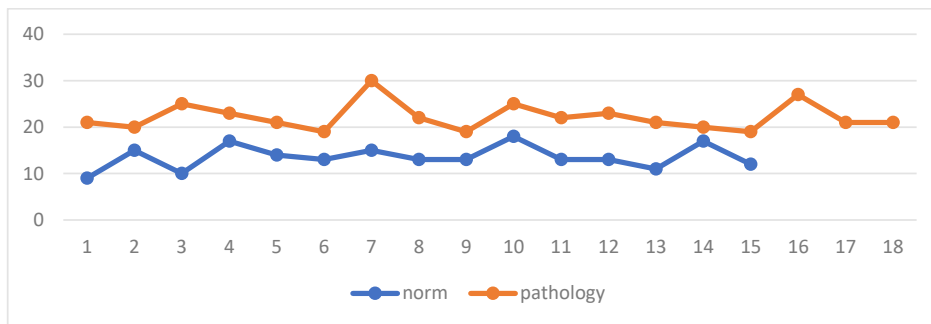
3-сурет. Кластерлеу нәтижесі (a), өкпенің пайызы (b)

Дерекқордан кездейсоқ таңдалған 33 сурет өңделді және оларға автокорреляция функциясы қолданылды (кесте -1).

Кесте - 1. Автокорреляциялық функция және k-орташа мәндер әдістерін қолдану нәтижесі мәндерінің кестесі

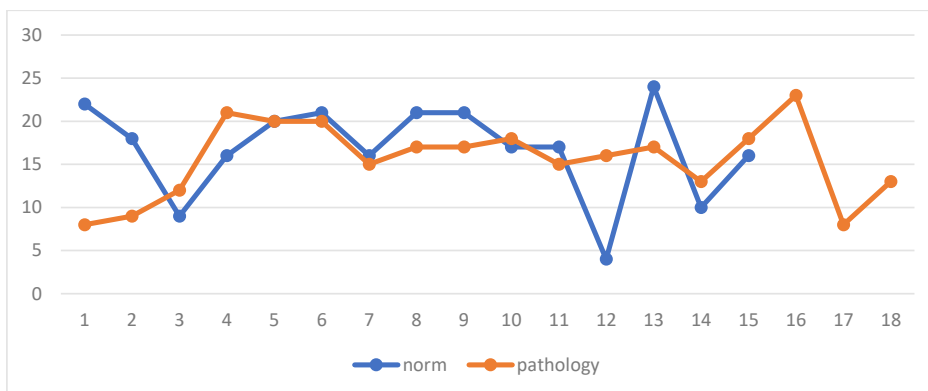
Суреттер атауы	Автокорреляциялық функция әдісі	K-means әдісі
Normal-57.png	9	22
person76_bacteria_371.jpeg	21	8
Normal-61.png	15	18
person77_bacteria_377.jpeg	20	9
Normal-62.png	10	9
person80_virus_150.jpeg	25	12
Normal-64.png	17	16
person82_virus_154.jpeg	23	21
Normal-65.png	14	20
person83_virus_156.jpeg	21	20
Normal-69.png	13	21
person88_virus_163.jpeg	19	20
Normal-72.png	15	16
person88_virus_165.jpeg	30	15
Normal-74.png	13	21
person89_virus_168.jpeg	22	17
Normal-76.png	13	21
person95_virus_177.jpeg	19	17
Normal-78.png	18	17
person96_virus_178.jpeg	25	18
Normal-80.png	13	17
Normal-81.png	13	4
person97_virus_181.jpeg	22	15
person98_virus_182.jpeg	23	16
person99_virus_183.jpeg	21	17
Normal-86.png	11	24
person100_virus_184.jpeg	20	13
Normal-87.png	17	10
Normal-91.png	12	16
person102_virus_189.jpeg	19	18
person105_virus_192.jpeg	27	23
person106_virus_194.jpeg	21	8
person106_virus_195.jpeg	21	13

Төменде кеуде қуысының қалыпты жағдайының ауытқу графигі және автокорреляциялық функциялар әдісін қолдану нәтижесінде патологияның пайда болуы келтірілген (4-сурет).



4-сурет. Автокорреляция функциясының көмегімен қалыпты және патологиялық кескіндердің көріну графигі

Паталогияны анықтауда автокорреляциялық функция әдісінің тиімділігін көру үшін ең танымал k-орташа әдісі (Поручиков, 2016) қарастырылды және оның графигі төменде көрсетілген (сурет-5).



5-сурет. K-орташа әдісін қолдану нәтижесі

Қорытынды. Рентгенограммаларды алдын ала сегменттеу алгоритмдері мен автокорреляциялық функцияларды қолдануға негізделген әдістер паталогияны тану дәлдігіне шамамен 98% жетуге мүмкіндік береді. Бұл дәлдік шағын жаттығулар жиынтығымен анықталады, сондықтан алдағы жұмыста рентгенограммалардың көбірек санына дифференциация жүргізу жоспарлануда және нейрондық желі реттеледі. Осылайша, болашақта бұл әдіс ауруларды диагностикалау процесін жылдамдатады және қайталанатын зерттеулердің үлесін азайтады. Сипатталған зерттеу әдісін тексеру үшін 33 флюорографиялық сурет таңдалды. Суреттердің бір жартысында пневмонияға байланысты әртүрлі паталогиялар болды, ал қалған жартысы сау өкпенің суреттері болды.

Рентгенологтар Қазақстан Республикасындағы денсаулық сақтау саласын цифрландырудың бірнеше өзекті мәселелерін шешуде шешуші рөл атқарады,

мысалы, оқыту үшін жоғары сапалы деректер жиынтығын құру, шешуді қажет ететін клиникалық мәселені анықтау және нәтижелерді түсіндіру. Рентгенолог тәжірибесінде жасанды интеллектті одан әрі енгізу үшін көптеген зерттеулер қажет, бірақ қазір автоматтандырылған медицина жүйесі жұмыс жүктемесінің бір бөлігін алып, дәрігердің жұмысын жеңілдетеді, сонымен қатар экономикалық жағдайды жақсартуға алады деп айта аламыз. денсаулық сақтаудың ресурстық базасының шығындарын азайту арқылы.

Information about the authors:

Shekerbek A.A. – doctoral student of the Department of Information Systems of the L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. E-mail: shekerbek80@mail.ru, <https://orcid.org/0000-0002-1088-4239>;

Abdikerimova G.B. – acting associate professor of the Department of Information Systems of L. N. Gumilyov Eurasian National University, PhD, Astana, Kazakhstan. E-mail: gulzira1981@mail.ru, h-index: 4 (Scopus) Scopus ID: 57203506578, <https://www.scopus.com/authid/detail.uri?authorId=57203506578>, <http://orcid.org/0000-0002-4953-0737>;

Sabyr A. M. – teacher of the Department of Information Technologies of the Kazakh University of technology and business, Astana, Kazakhstan. E-mail: aigerraa01@gmail.com;

Abylkair Zh.S. – teacher of the Department of Information Technologies of the Kazakh University of technology and business, Astana, Kazakhstan. E-mail: zhain1504@gmail.com.

ӘДЕБИЕТТЕР:

Поручиков М.А., Анализ данных: учеб. пособие / М.А. Поручиков. – Самара: Изд-во Самарского университета, 2016. – 88 с. (орыс тілінде).

Abdikerimova G.B. et al. Software tools for cell walls segmentation in microphotography // Journal of Theoretical and Applied Information Technology. – 2018. – Т. 96. – №. 15. – С. 4783-4793. (ағылшын тілінде).

Driver C.N., Bowles B.S., Bartholmai B.J. & Greenberg-Worisek A.J. (2020). Artificial Intelligence in Radiology: A Call for Thoughtful Application. Clinical and translational science, 13(2), 216–218. <https://doi.org/10.1111/cts.12704>. (ағылшын тілінде).

Haralick R.M. (1979). Statistical and structural approaches to texture. Proceedings of the IEEE, 67(5), 786-804. <https://doi.org/10.1109/PROC.1979.11328>, (ағылшын тілінде).

<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>, (ағылшын тілінде).

Pisaniello H.L. & Dixon W.G. (2020). What does digitalization hold for the

creation of real-world evidence?. *Rheumatology* (Oxford, England), 59(1), 39–45. <https://doi.org/10.1093/rheumatology/kez068>, (ағылшын тілінде).

Shekerbek A. et al. Application of mathematical methods and machine learning algorithms for classification of X-ray images//*Eastern-European Journal of Enterprise Technologies*. – 2022. Т. 3. – №.2. – С. 6–17, <https://doi.org/10.15587/1729-4061.2022.259710>, (ағылшын тілінде).

Taubayev G. et al. Machine learning algorithms and classification of textures // *Journal of Theoretical and Applied Information Technology*. – 2020. – Т. 98. – №. 23. – С. 3854-3866, (ағылшын тілінде).

REFERENCES:

M.A. Poruchikov, data analysis: a study. the benefits / Ma lieutenants. - Samara: edition of the University of Samara, 2016-88 P. (into Russian).

Abdikerimova G.B. et al. software tools for cell wall segmentation in microphotography // *Journal of theoretical and Applied Information Technologies*. - 2018. - Vol. 96. -№. 15. - pp. 4783-4793. (in English).

The Driver K.N., Bowles B.S., Bartholmay B.J., and Greenberg-Warisek A.J. (2020). Artificial intelligence in radiology: a call for thoughtful use. *Clinical and Translational Science*, 13 (2), 216-218. <https://doi.org/10.1111/cts.12704>. (in English).

Haralik R.M. (1979). Statistical and structural approaches to texture. *IEEE proceedings*, 67 (5), 786-804. <https://doi.org/10.1109/PROC.1979.11328>, (in English).

<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>, (in English).

Pisaniello H.L., and Dixon W.G. (2020). What does digitalization provide to create concrete evidence?. *Rheumatology*(Oxford, England), 59 (1), 39-45. <https://doi.org/10.1093/rheumatology/kez068> , (in English).

The use of mathematical methods and algorithms of machine learning for the classification of X-ray images // *Eastern European Journal of Corporate Technologies*. - 2022. Vol – 3. -№.2. - pp. 6-17, <https://doi.org/10.15587/1729-4061.2022.259710> , (in English).

Machine learning algorithms and classification of textures / / *Journal of theoretical and Applied Information Technologies*. - 2020. - Vol. 98. -№. 23. - pp. 3854-3866, (in English).

МАЗМҰНЫ

А.С. Баймаханова, А.Ж. Сейтмуратов DEEP LEARNING АЛГОРИТМІН ҚОЛДАНУ НЕГІЗІНДЕ ЦИФРЛЫҚ ҚҰЖАТТАРДЫ ЖІКТЕУ.....	5
М.А. Болатбек, Ш.Ж. Мусиралиева, К. Багитова, А.Т. Нюсупов, Е. Абайұлы ВЕБ-РЕСУРСТАРДАҒЫ ФИШИНГТІК ХАБАРЛАМАЛАР ЖӘНЕ ОЛАРДЫ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ АРҚЫЛЫ АНЫҚТАУ.....	16
М.А. Кантуреева, А.Ш. Хасенов, Д.А. Тусупов, А.Б. Закирова, А.З. Алимагамбетова ЭВАКУАЦИЯ ДИНАМИКАСЫНА АРНАЛҒАН FLOOR FIELD МОДЕЛІ...30	30
А.Д. Кубегенова, К.Т. Искаков, Е.С. Кубегенов, О.И. Криворотько ДЕРЕКТЕРДІ ИНТЕЛЕКТУАЛДЫ ТАЛДАУ АРҚЫЛЫ ЭПИДЕМИОЛОГИЯЛЫҚ ЖАҒДАЙДЫ БАҚЫЛАУ ЖӘНЕ МОДЕЛЬДЕУ.....	43
Г. Қалман, М.А. Самбетбаева, Д.А. Ақтаева, А.С. Илюбаев МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН АНАФОРАНЫ ШЕШУ МОДЕЛІ.....	56
С.Т. Мамбетов, Е.Е. Бегимбаева, С.К. Джолдасбаев, Б.О. Куламбаев, Г.Н. Казбекова АҚПАРАТТЫҚ ЖҮЙЕНІҢ ҚАУІПТЕРІ МЕН ОСАЛ ТҰСТАРЫНЫҢ МОНИТОРИНГІ ТУРАЛЫ.....	68
У.Т. Махажанова, Б. Тасуов, А.А. Муханова, А. Мухиядин, Р.К. Жеткиншеков БҰЛДЫР ЖИЫНДАР ТЕОРИЯСЫ НЕГІЗІНДЕ БИЗНЕСТІҢ НЕСИЕ ҚАБІЛЕТІЛІГІН БАҒАЛАУ АЛГОРИТМІ.....	81
Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали, Г.Б. Турмуханова СУ ЭКО ЖҮЙЕЛЕРІН ЗЕРТТЕУДЕ АТЖ ӨЗІРЛЕУГЕ ҚОЙЫЛАТЫН ТАЛАПТАР.....	93
А.А. Муханова, У.Т. Махажанова, Н.Д. Мархабатов, Б. Тасуов, Ж.Б. Ламашева ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ТАЛДАУДА БҰЛДЫР ЛОГИКАНЫ ҚОЛДАНУ.....	106

Н.А. Сейлова, А.Б. Батыргалиев, Ж.А. Джангозин, Д.А. Байбатчаева, Н. Нұрғабылов ШУ КЕДЕЛДЕРІН БҮРКЕУДІҢ САПАСЫН БАҒАЛАУ ӘДІСТЕМЕСІ.....	120
А.Ш. Хасенов, М.А. Кантурсева, Д.А. Тусупов, А.С. Омарбекова, Г.Б. Абдикеримова АГЕНТТІК МОДЕЛЬДЕУ ЖҮЙЕСІНДЕ ЭВАКУАЦИЯ МОДЕЛІН ЖҮЗЕГЕ АСЫРУ ТӘСІЛІ.....	134
А. Шаушенова, А. Нурпейсова, Д. Досалянов, Г. Мауина ПРОКТОРИНГ ЖҮЙЕСІНДЕ ЖАСАНДЫ НЕЙРОНДЫҚ ЖЕЛІЛЕРГЕ НЕГІЗДЕЛГЕН СӨЙЛЕУДІ ТАҢУ МӘСЕЛЕЛЕРІ.....	146
А.Ә. Шекербек, Г.Б. Абдикеримова, Ә.М. Сабыр, Ж.С. Әбілқайыр КЕУДЕ КЛЕТКАСЫНЫҢ ПАТОЛОГИЯСЫН АНЫҚТАУ ҮШІН ӘДІС ПЕН АЛГОРИТМДІ ҚОЛДАНУ.....	159

СОДЕРЖАНИЕ

А.С. Баймаханова, А.Ж. Сейтмуратов КЛАССИФИКАЦИЯ ЦИФРОВЫХ ДОКУМЕНТОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ АЛГОРИТМА DEEP LEARNING.....	5
М.А. Болатбек, Ш.Ж. Мусиралиева, К. Багитова, А.Т. Нюсупов, Е. Абайулы ФИШИНГОВЫЕ СООБЩЕНИЯ НА ВЕБ-РЕСУРСАХ И ИХ ОПРЕДЕЛЕНИЕ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ.....	16
М.А. Кантуреева, А.Ш. Хасенов, Д.А. Тусупов, А.Б. Закирова, А.З. Алимагамбетова FLOOR FIELD МОДЕЛЬ ДЛЯ ДИНАМИКИ ЭВАКУАЦИИ.....	30
А.Д. Кубегенова, К.Т. Искаков, Е.С. Кубегенов, О.И. Криворотько МОНИТОРИНГ И МОДЕЛИРОВАНИЕ ЭПИДЕМИОЛОГИЧЕСКОЙ СИТУАЦИИ С ПОМОЩЬЮ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ.....	43
Г. Қалман, М.А. Самбетбаева, Д.А. Актаева, А.С. Илюбаев МОДЕЛЬ РАЗРЕШЕНИЯ АНАФОРЫ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ.....	56
С.Т. Мамбетов, Е.Е. Бегимбаева, С.К. Джолдасбаев, Б.О. Куламбаев, Г.Н. Казбекова О МОНИТОРИНГЕ УГРОЗ И УЯЗВИМОСТЕЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ.....	68
У.Т. Махажанова, Б. Тасуов, А.А. Муханова, А. Мухиядин, Р.К. Жеткиншеков АЛГОРИТМ ОЦЕНКИ КРЕДИТОСПОСОБНОСТИ БИЗНЕСА НА ОСНОВЕ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ.....	81
Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали, Г.Б. Турмуханова ТРЕБОВАНИЯ К РАЗРАБОТКЕ ИАС-ИССЛЕДОВАНИЙ ВОДНЫХ ЭКОСИСТЕМ.....	93
А.А. Муханова, У.Т. Махажанова, Н.Д. Мархабатов, Б. Тасуов, Ж.Б. Ламашева ПРИМЕНЕНИЕ НЕЧЕТКОЙ ЛОГИКИ ПРИ АНАЛИЗЕ ЭКОНОМИЧЕСКИХ СИСТЕМ.....	106

Н.А. Сейлова, А.Б. Батыргалиев, Ж.А. Джангозин, Д.А. Байбатчаева, Н. Нұрғабылов МЕТОДИКА ОЦЕНКИ КАЧЕСТВА МАСКИРУЮЩИХ ШУМОВЫХ ПОМЕХ.....	120
А.Ш. Хасенов, М.А. Кантуреева, Д.А. Тусупов, А.С. Омарбекова, Г.Б. Абдикеримова ПОДХОД К РЕАЛИЗАЦИИ МОДЕЛИ ЭВАКУАЦИИ В СИСТЕМЕ АГЕНТНОГО МОДЕЛИРОВАНИЯ.....	134
А.Г. Шаушенова, А.А. Нурпейсова, Д.Б. Досалянов, Г.М. Мауина ПРОБЛЕМЫ РАСПОЗНАВАНИЯ РЕЧИ НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В СИСТЕМЕ ПРОКТОРИНГА.....	146
А.А. Шекербек, Г.Б. Абдикеримова, А.М. Сабыр, Ж.С. Абулхаир ПРИМЕНЕНИЕ МЕТОДА И АЛГОРИТМА ДЛЯ ВЫЯВЛЕНИЯ ПАТОЛОГИИ ГРУДНОЙ КЛЕТКИ.....	159

CONTENTS

A. Baimakhanova, A. Seitmuratov CLASSIFICATION OF DIGITAL DOCUMENTS USING DEEP LEARNING ALGORITHM.....	5
M. Bolatbek, Sh. Musiralieva, K Bagitova, A. Нюсупов, E. Abaiuly PHISHING MESSAGES ON WEB RESOURCES AND THEIR DETECTION BY MACHINE LEARNING METHODS.....	16
M. Kantureyeva, A. Khassenov, D. Tussupov, A. Zakirova, A. Alimagambetova FLOOR FIELD MODEL FOR EVACUATION DYNAMICS.....	30
A.D. Kubegenova, K.T. Iskakov, E.S. Kubegenov, O.I. Krivorotko MONITORING AND MODELING OF THE EPIDEMIOLOGICAL SITUATION USING DATA MINING.....	43
G. Kalman, M.A. Sambetbayeva, A.C. Ilyubayev, D.A. Aktaeva ANAPHORA RESOLUTION MODEL BASED ON MACHINE LEARNING METHODS.....	56
S.T. Mambetov, Ye.Ye. Begimbayeva, S. Joldasbayev, B.O. Kulambayev, G.N. Kazbekova ABOUT MONITORING THREATS AND VULNERABILITIES OF THE INFORMATION SYSTEM.....	68
U. Makhazhanova, B. Tassuov, A. Mukhanova, A. Mukhiyadin, R. Zetkinshekov AN ALGORITHM FOR ASSESSING THE CREDITWORTHINESS OF A BUSINESS BASED ON THE THEORY OF FUZZY SETS.....	81
R.M. Moldasheva, A.A. Ismailova, A.K. Zhamangara, A.M. Zadagali, G.B. Turmukhanova REQUIREMENTS TO DEVELOPMENT OF IAS FOR RESEARCH OF AQUEOUS ECOSYSTEMS.....	93
A. Mukhanova, U. Makhazhanova, N. Markhabatov, B. Tassuov, Zh. Lamasheva APPLICATION OF FUZZY LOGIC IN THE ANALYSIS OF ECONOMIC SYSTEMS N.....	106

N.A. Seilova, A. Batyrgaliyev, Zh. Dzhangozin, D. Baibatchayeva, N. Nurgabylov METHOD FOR ASSESSING THE QUALITY OF MASKING NOISE INTERFERENCES.....	120
A. Khassenov, M. Kantureyeva, D. Tussupov, A. Omarbekova, G. Abdikerimova APPROACH TO THE IMPLEMENTATION OF EVACUATION MODEL IN THE AGENT-BASED MODELING SYSTEM.....	134
A.G. Shaushenova, A.A. Nurpeisova, D.B. Dosalyanov, G.M. Mauina PROBLEMS OF SPEECH RECOGNITION BASED ON ARTIFICIAL NEURAL NETWORKS IN THE PROCTORING SYSTEM.....	146
A. Shekerbek, G. Abdikerimova, A. Sabyr, Zh. Abilkaiyr APPLICATION OF THE METHOD AND ALGORITHM FOR THE DETECTION OF CHEST PATHOLOGY.....	159

**Publication Ethics and Publication Malpractice
the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

www.nauka-nanrk.kz

<http://physics-mathematics.kz/index.php/en/archive>

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*

Заместитель директор отдела издания научных журналов НАН РК *Р. Жәліқызы*

Редакторы: *М.С. Ахметова, Д.С. Аленов*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 15.09.2022.

Формат 60x88/8. Бумага офсетная. Печать – ризограф.

10,5 п.л. Тираж 300. Заказ 4.