

ISSN 2616-7182
eISSN 2663-1326

Л.Н. Гумилев атындағы Еуразия ұлттық университетінің
ХАБАРШЫСЫ

BULLETIN
of L.N. Gumilyov Eurasian
National University

ВЕСТНИК
Евразийского национального
университета имени Л.Н. Гумилева

МАТЕМАТИКА. КОМПЬЮТЕРЛІК ҒЫЛЫМДАР. МЕХАНИКА сериясы

MATHEMATICS. COMPUTER SCIENCE. MECHANICS Series

Серия **МАТЕМАТИКА. КОМПЬЮТЕРНЫЕ НАУКИ. МЕХАНИКА**

№4(145)/2023

1995 жылдан бастап шығады

Founded in 1995

Издается с 1995 года

Жылына 4 рет шығады

Published 4 times a year

Выходит 4 раза в год

Астана, 2023
Astana, 2023

БАС РЕДАКТОРЫ

Темірғалиев Н., ф.-м.ғ.д., проф., Л.Н. Гумилев ат. ЕҰУ, Астана, Қазақстан

Бас редактордың орынбасары

Жұбанышева А.Ж.

PhD, Л.Н. Гумилев ат. ЕҰУ, Астана, Қазақстан

Бас редактордың орынбасары

Наурызбаев Н.Ж.

PhD, Л.Н. Гумилев ат. ЕҰУ, Астана, Қазақстан

Редакция алқасы

Абакумов Е.В.

PhD, проф., Париж-Эст университеті, Марн-Ла-Вале, Париж, Франция

Алексеева Л.А.

ф.-м.ғ.д., проф., ҚР БҰҒМ Математика және математикалық модельдеу институты, Алматы, Қазақстан

Алимхан Қилан

PhD, проф., Л.Н. Гумилев ат. ЕҰУ, Астана, Қазақстан

Балтаева У.

ф.-м.ғ.д., Мамун Хорезм академиясы, Хорезм, Өзбекстан

Бекенов М.И.

ф.-м.ғ.к., доцент, Л.Н. Гумилев ат. ЕҰУ, Астана, Қазақстан

Гогинава У.

ф.-м.ғ.д., проф., Ив. Джавахишвили Тбилиси мемлекеттік университеті, Тбилиси, Грузия

Голубов Б.И.

ф.-м.ғ.д., проф., Мәскеу физика-техника институты (мемлекеттік университет) Долгопрудный, Ресей

Зунг Динь

ф.-м.ғ.д., проф., Информатикалық технологиялар институты, Вьетнам ұлттық университеті, Ханой, Вьетнам

Иванов В.И.

ф.-м.ғ.д., проф., Тула мемлекеттік университеті, Тула, Ресей

Иосевич А.

PhD, проф., Рочестер университеті, Нью-Йорк, АҚШ

Кобельков Г.М.

ф.-м.ғ.д., проф., М.В. Ломоносов атындағы Мәскеу мемлекеттік университеті, Мәскеу, Ресей

Курина Г.А.

ф.-м.ғ.д., проф., Воронеж мемлекеттік университеті, Воронеж, Ресей

Марков В.В.

ф.-м.ғ.д., проф., РҒА В.А. Стеклов атындағы Мәскеу мемлекеттік институты, Мәскеу, Ресей

Мейрманов А.М.

ф.-м.ғ.д., проф., Байланыс және информатика Мәскеу техникалық университеті, Мәскеу, Ресей

Омарбекова А.С.

т.ғ.к., Л.Н. Гумилев ат. ЕҰУ, Астана, Қазақстан

Смелянский Р.Л.

ф.-м.ғ.д., проф., М.В. Ломоносов атындағы Мәскеу мемлекеттік университеті, Мәскеу, Ресей

Умирбаев У.У.

ф.-м.ғ.д., проф., Уейна мемлекеттік университеті, Детройт, АҚШ

Холщевникова Н.Н.

ф.-м.ғ.д., проф., "Станкин" Мәскеу мемлекеттік техникалық университеті, Мәскеу, Ресей

Шмайссер Ханс-Юрген

Хабилит. докторы, проф., Фридрих-Шиллер университеті, Йена, Германия

Редакцияның мекенжайы: 010008, Қазақстан, Астана қ., Сәтпаев к-сі, 2, 402 бөлме.

Тел: +7 (7172) 709-500 (ішкі 31-410). E-mail: vest_math@enu.kz

Жауапты редактор: А.Ж. Жұбанышева

Л.Н. Гумилев атындағы Еуразия ұлттық университетінің хабаршысы.

МАТЕМАТИКА. КОМПЬЮТЕРЛІК ҒЫЛЫМДАР. МЕХАНИКА сериясы

Меншіктенуші: Л.Н. Гумилев атындағы Еуразия ұлттық университеті.

Мерзімділігі: жылына 4 рет.

Қазақстан Республикасы Ақпарат және қоғамдық даму министрлігімен тіркелген. 02.02.2021 ж.

№ KZ65VPY00031936 қайта есепке қою туралы куәлігі.

Типографияның мекенжайы: 010008, Қазақстан, Астана қ., Қажымұқан к-сі, 12/1,

тел: +7 (7172)709-500 (ішкі 31-410).

EDITOR-IN-CHIEF

Nurlan Temirgaliyev

Prof., Doctor of Phys.-Math. Sciences, L.N.Gumilyov ENU, Astana, Kazakhstan

Deputy Editor-in-Chief

Aksaule Zhubanysheva

PhD, L.N.Gumilyov ENU, Astana, Kazakhstan

Deputy Editor-in-Chief

Nurlan Nauryzbayev

PhD, L.N.Gumilyov ENU, Astana, Kazakhstan

Editorial board:

Evgueni Abakumov

*PhD, Prof., University Paris-Est, Marne-la-Vallee
Paris, France*

Lyudmila Alexeyeva

*Doctor of Phys.-Math. Sci., Prof., Institute of Mathematics and Mathematical Modeling Ministry of Education
and Science Republic of Kazakhstan, Almaty, Kazakhstan*

Alexander Iosevich

PhD, Prof., University of Rochester, New York, USA

Alimhan Keylan

PhD, Prof., L.N. Gumilyov ENU, Astana, Kazakhstan

Umida Baltaeva

*Doctor of Phys.-Math. Sci., Khorezm Mamun Academy, Khorezm,
Uzbekistan*

Makhsut Bekenov

*Candidate of Phys.-Math. Sci., Assoc.Prof.
L.N. Gumilyov ENU, Astana, Kazakhstan*

Ushangi Goginava

*Doctor of Phys.-Math. Sci., Prof.
Iv. Javakhishvili Tbilisi State University, Tbilisi, Georgia*

Boris Golubov

*Doctor of Phys.-Math. Sci., Prof., Moscow Institute of Physics and
Technology (State University)
Dolgoprudnyi, Russia*

Dũng Dinh

*Doctor of Phys.-Math. Sci., Prof., Information Technology Institute,
Vietnam National University, Hanoi, Vietnam*

Valerii Ivanov

Doctor of Phys.-Math. Sci., Prof., Tula State University, Tula, Russia

Georgii Kobel'kov

*Doctor of Phys.-Math. Sci., Prof., Lomonosov Moscow State University,
Moscow, Russia*

Galina Kurina

*Doctor of Phys.-Math. Sci., Prof., Voronezh State University, Voronezh,
Russia*

Vladimir Markov

*Doctor of Phys.-Math. Sci., Prof., Steklov Mathematical
Institute of Russian Academy of Sciences, Moscow, Russia*

Anvarbek Meirmanov

*Doctor of Phys.-Math. Sci., Prof., Moscow Technical University of Com-
munications and Informatics, Moscow, Russia*

Asel Omarbekova

Cand. of Tech. Sci., L.N. Gumilyov ENU, Astana, Kazakhstan

Ruslan Smelyansky

*Doctor of Phys.-Math. Sci., Prof., Lomonosov Moscow State University,
Moscow, Russia*

Ualbay Umirbaev

*Doctor of Phys.-Math. Sci., Prof.,
Wayne State University, Detroit, USA*

Natalya Kholshchevnikova

*Doctor of Phys.-Math. Sci., Prof., Moscow State
Technological University "Stankin", Moscow, Russia*

Hans-Juergen Schmeisser

*Dr. habil., Prof., Friedrich-Shiller University
Jena, Germany*

Editorial address: 2, Satpayev str., of. 402, Astana, Kazakhstan, 010008.

Tel.: +7 (7172) 709-500 (ext. 31-410). E-mail: vest_math@enu.kz

Responsible Editor-in-Chief: Aksaule Zhubanysheva

Bulletin of the L.N. Gumilyov Eurasian National University.

MATHEMATICS. COMPUTER SCIENCE. MECHANICS Series

Owner: L.N. Gumilyov Eurasian National University. Periodicity: 4 times a year.

Registered by the Ministry of Information and Social Development of the Republic of Kazakhstan. Rediscount certificate № KZ65VPY00031936 dated 02.02.2021.

Address of printing house: 12/1 Kazhimukan str., Astana, Kazakhstan 010008; tel: +7 (7172) 709-500 (ext.31-410).

© L.N. Gumilyov Eurasian National University

ГЛАВНЫЙ РЕДАКТОР

Темиргалиев Н., д.ф.-м.н., проф., ЕНУ имени Л.Н.Гумилева, Астана, Казахстан

Зам. главного редактора

Жубанышева А.Ж.

PhD, ЕНУ имени Л.Н.Гумилева, Астана, Казахстан

Зам. главного редактора

Наурызбаев Н.Ж.

PhD, ЕНУ имени Л.Н.Гумилева, Астана, Казахстан

Редакционная коллегия

Абакумов Е.В.

PhD, проф., Университет Париж-Эст, Марн-Ла-Вале, Париж, Франция

Алексеева Л.А.

д.ф.-м.н., проф., Институт математики и математического моделирования МОН РК, Алматы, Казахстан

Алимхан Килян

PhD, проф., ЕНУ имени Л.Н.Гумилева, Астана, Казахстан

Балтаева У.

д.ф.-м.н., Хорезмская академия Маъмуна, Хорезм, Узбекистан

Гогинава У.

д.ф.-м.н., проф., Тбилисский государственный университет имени Ив. Джавахишвили, Тбилиси, Грузия

Голубов Б.И.

д.ф.-м.н., проф., Московский физико-технический институт (государственный университет), Долгопрудный, Россия

Зунг Динь

д.ф.-м.н., проф., Институт информационных технологий, Вьетнамский национальный университет, Ханой, Вьетнам

Иванов В.И.

д.ф.-м.н., проф., Тульский государственный университет, Тула, Россия

Иосевич А.

PhD, проф., Рочестерский университет, Нью-Йорк, США

Кобельков Г.М.

д.ф.-м.н., проф., МГУ имени М.В. Ломоносова, Москва, Россия

Курина Г.А.

д.ф.-м.н., проф., Воронежский государственный университет, Воронеж, Россия

Марков В.В.

д.ф.-м.н., проф., Математический институт им. В.А. Стеклова РАН, Москва, Россия

Мейрманов А.М.

д.ф.-м.н., проф., Московский технический университет связи и информатики, Москва, Россия

Омарбекова А.С.

к.т.н., ЕНУ имени Л.Н.Гумилева, Астана, Казахстан

Смелянский Р.Л.

д.ф.-м.н., проф., МГУ имени М.В. Ломоносова, Москва, Россия

Умирбаев У.У.

д.ф.-м.н., проф., Государственный университет Уейна, Детройт, США

Холщевникова Н.Н.

д.ф.-м.н., проф., Московский государственный технологический университет "Станкин", Москва, Россия

Шмайссер Ханс-Юрген

Хабилит. доктор, проф., Университет Фридрих-Шиллера, Йена, Германия

Адрес редакции: 010008, Казахстан, г. Астана, ул. Сатпаева, 2, каб. 402

Тел: +7 (7172) 709-500 (вн. 31-410). E-mail: vest_math@enu.kz

Ответственный редактор: А.Ж. Жубанышева

Вестник Евразийского национального университета имени Л.Н. Гумилева.

Серия МАТЕМАТИКА. КОМПЬЮТЕРНЫЕ НАУКИ. МЕХАНИКА

Собственник: Евразийский национальный университет имени Л.Н. Гумилева.

Периодичность: 4 раза в год.

Зарегистрировано Министерством информации и общественного развития Республики Казахстан.

Свидетельство о постановке на переучет № KZ65VPY00031936 от 02.02.2021 г.

Адрес типографии: 010008, Казахстан, г. Астана, ул. Кажымукана, 12/1, тел.: +7 (7172)709-500 (вн.31-410).

Л.Н. Гумилев атындағы Еуразия ұлттық университетінің хабаршысы.
Математика. Компьютерлік ғылымдар. Механика сериясы, №4(145)/2023

Bulletin of L.N. Gumilyov Eurasian National University.
Mathematics. Computer science. Mechanics series, №4(145)/2023

Вестник Евразийского национального университета имени Л.Н.Гумилева.
Серия Математика. Компьютерные науки. Механика, №4(145)/2023

МАЗМҰНЫ
CONTENTS
СОДЕРЖАНИЕ

- Утесов А.Б.* Жалпыланған Соболев кластары функцияларын қалыптастыру есебінің толық $K(E)D$ -зерттеуі
Utesov A.B. Full $C(N)D$ -research of the problem of recovery functions from the generalized Sobolev class
Утесов А.Б. Полное $K(B)D$ -исследование задачи восстановления функций из обобщенного класса Соболева 6
- Кожирбаев Ж.М.* Қазақ тілін тану жүйесін өз бетінше оқыту
Kozhirbayev Zh.M. Self-Supervised Training for the Kazakh Speech Recognition System
Кожирбаев Ж.М. Самообучение для улучшения системы распознавания речи на казахском языке 12
- Әбікенова Ш.К., Марчеллони Р., Әйтимова Ш.Т.* Кәсіби тәуекелдің статистикалық көрсеткіштер негізінде жіктелуінің математикалық аспектілері
Abikenova Sh., Marcelloni R., Aitimova Sh. Mathematical aspects of occupational risk and its classification considering statistical indicators
Абикенова Ш.К., Марчеллони Р., Айтимова Ш.Т. Математические аспекты профессионального риска и его классификации с учетом статистических показателей 23

МРНТИ: 27.25.19

А.Б. Утесов

Актюбинский региональный университет имени К. Жубанова, пр. А.Молдагуловой, 34,
Актобе, 030000, Казахстан
(E-mail: adilzhan_71@mail.ru)

Полное К(В)П-исследование задачи восстановления функций из обобщенного класса Соболева

Аннотация: В данной работе проведено полное К(В)П-исследование задачи восстановления функций из обобщенного класса Соболева $W_2^{\omega_r}$ в случае, когда числовая информация объема N о восстанавливаемой функции f снимается с линейных функционалов. Именно, во-первых, в метрике L^q ($2 \leq q \leq \infty$) установлен точный порядок погрешности восстановления функций из классов $W_2^{\omega_r}$. Во-вторых, предложен конкретный вычислительный агрегат, реализующий точный порядок и найдена его предельная погрешность $\bar{\varepsilon}_N$, сохраняющая точный порядок и неуплучшаемая по порядку. В-третьих, доказано, что любой вычислительный агрегат, построенный по коэффициентам Фурье восстанавливаемой функции не имеет предельной погрешности, лучшей (по порядку) чем $\bar{\varepsilon}_N$.

Ключевые слова: К(В)П-исследование, линейный функционал, вычислительный агрегат, точный порядок погрешности восстановления, предельная погрешность, обобщенный класс Соболева.

DOI: <https://doi.org/10.32523/2616-7182/bulmathenu.2023/4.1>

2000 Mathematics Subject Classification: 41A99

1. Определение обобщенного класса Соболева. Для заданного числа $r > 0$ всякая непрерывная неубывающая на $[0, 1]$ функция $\omega_r(\delta)$ называется функцией типа модуля гладкости порядка r , если $\omega_r(0) = 0$ и существует величина $C_1(r) > 0$ такая, что

$$\frac{\omega_r(\mu)}{\mu^r} \leq C_1(r) \frac{\omega_r(\delta)}{\delta^r}$$

для всех $0 < \delta < \mu \leq 1$.

В качестве функций типа модуля гладкости порядка r можно указать функции

$$\omega_r^{(1)}(\delta) = \begin{cases} \delta^r \ln^{r_1}(2/\delta), & \text{если } \delta \in (0, 1]; \\ 0, & \text{если } \delta = 0 \end{cases}$$

и

$$\omega_r^{(2)}(\delta) = \begin{cases} \delta^r \ln^{r_1}(2/\delta) \ln^{r_2}(\ln \ln(2/\delta)), & \text{если } \delta \in (0, 1]; \\ 0, & \text{если } \delta = 0 \end{cases}$$

с параметрами $r > 0, -\infty < r_1 < +\infty$ и $r > 0, r_1 > 0, -\infty < r_2 < +\infty$ соответственно.

Обобщенный класс Соболева $W_2^{\omega_r} \equiv W_2^{\omega_r}(0, 1)^s$, по определению, состоит из всех суммируемых 1-периодических по каждой переменной функций $f(x) = f(x_1, \dots, x_s)$,

удовлетворяющих условию

$$\sum_{m \in Z^s} |\hat{f}(m)|^2 (\omega_r^{-2}(1/\bar{m}_1) + \dots + \omega_r^{-2}(1/\bar{m}_s)) \leq 1,$$

где $m = (m_1, \dots, m_s) \in Z^s$, $\hat{f}(m) = \int_{[0,1]^s} f(x) e^{-2\pi i(m,x)} dx$ – тригонометрические коэффициенты Фурье-Лебега функции f , $(m, x) = m_1 x_1 + \dots + m_s x_s$, $\bar{m}_j = \max\{1; |m_j|\}$ для каждого $j = 1, \dots, s$.

Приведенный выше класс является более тонкой шкалой классификации периодических функций многих переменных в зависимости от скорости убывания их коэффициентов Фурье: при $\omega_r(\delta) = \delta^r$ класс $W_2^{\omega_r}(0, 1)^s$ обращается в класс Соболева $W_2^r(0, 1)^s$.

Класс $W_2^{\omega_r}$ впервые был рассмотрен в [1] при изучении задачи интегрирования функций на функциональных классах. Затем в [2] в гильбертовой метрике были найдены точные порядки погрешностей, возникающих при восстановлении функций из рассматриваемых классов вычислительными агрегатами, построенными по их значениям в конечном числе точек. Там же, аналогичные результаты были получены при дискретизации решений уравнения теплопроводности с начальными условиями из классов $W_2^{\omega_r}$. Отметим также [3], посвященной к задаче дискретизации классических решений волнового уравнения с начальными условиями f_1 и f_2 из обобщенных классов Соболева.

В данной работе, следуя [4-6], проведено полное К(В)П - исследование задачи восстановления функций из обобщенного класса $W_2^{\omega_r}$, т.е. последовательно решены задачи К(В)П-1 (восстановление по точной информации, в зависимости от вида функционалов и алгоритмов переработки полученной от них числовой информации, включает в себя всю известную теорию приближений, численный анализ, вычислительную математику, теорию функций - ряды Фурье, базисы), К(В)П-2 (в оптимальном вычислительном агрегате значения информационных функционалов можно заменить на близкие им значения с сохранением оптимальности, поиск наибольших таких расхождений образует самостоятельную задачу нахождения предельных погрешностей – новая оптимизационная задача) и К(В)П-3 (существует или не существует другой вычислительный агрегат со структурой, аналогичной структуре рассматриваемого оптимального вычислительного агрегата, и даже, быть может, более общей, но с большей по порядку предельной погрешностью).

2. Вспомогательные утверждения. При доказательстве приведенной ниже теоремы используется функция

$$\bar{\omega}_r(\delta) = \frac{1}{\omega_r(1) + 1} (\omega_r(\delta) + \delta), \delta \in [0, 1].$$

Эта функция строго возрастает на $[0, 1]$ и удовлетворяет равенству $\bar{\omega}_r(1) = 1$. Символом $\bar{\omega}_r^*(\delta)$ будем обозначать функцию, обратную к функции $\bar{\omega}_r$.

Непосредственно из определений функций ω_r и $\bar{\omega}_r$ вытекают следующие их свойства:

I. Пусть $a > 1$. Тогда для всех $\delta \in (0, 1/a]$ будет выполнено неравенство

$$\omega_r(a\delta) \leq C_1(r) a^r \omega_r(\delta);$$

II. При каждом $a(0 < a < 1)$ для всех $\delta \in (0, 1]$ справедливо неравенство

$$\omega_r(a\delta) \geq \frac{a^r}{C_1(r)} \omega_r(\delta);$$

III. Для всех $0 \leq \delta \leq 1$ имеет место неравенство

$$\omega_r(\delta) \geq \frac{\omega_r(1)}{C_1(r) + 1} \delta^r;$$

IV. Для некоторых величин $C_2(r) > 0$ и $C_3(r) > 0$ справедливы неравенства

$$C_2(r) \omega_r(\delta) \leq \bar{\omega}_r(\delta) \leq C_3(r) \omega_r(\delta), \delta \in [0, 1].$$

Лемма 1. Пусть для некоторого $C_4(r) > 0$ и для всех $0 \leq \delta, \eta \leq 1$ выполняется неравенство $\omega_r(\delta\eta) \leq C_4(r)\omega_r(\delta)\omega_r(\eta)$. Тогда существует $C_5(r) > 0$ такая, что

$$\bar{\omega}_r(\delta\eta) \leq C_5(r)\bar{\omega}_r(\delta)\bar{\omega}_r(\eta)$$

для всех $0 \leq \delta, \eta \leq 1$ и

$$\bar{\omega}_r^*(C_5(r)\delta\eta) \geq \bar{\omega}_r^*(\delta)\bar{\omega}_r^*(\eta)$$

для всех $0 \leq \delta, \eta \leq 1$, удовлетворяющих неравенству $C_5(r)\delta\eta \leq 1$.

Лемма 2. Существует величина $C_6(r) \leq 1$ такая, что $\bar{\omega}_r(C_6(r)\delta) \leq \delta^{1/r}$, $\delta \in [0, 1]$.

Лемма 3. Пусть $\|m\| = \max\{|m_1|, \dots, |m_s|\}$ для всех $m = (m_1, \dots, m_s)$ и $I_\tau = \{m \in Z^s : [2^{\tau-1}] \leq \|m\| < 2^\tau\}$, $\tau \in \{0, 1, 2, \dots\}$. Тогда существует величина $C_7(r) > 0$ такая, что для любой функции $f \in W_2^{\omega_r}$ справедливо неравенство

$$\sup_{\tau=0,1,2,\dots} \omega_r^{-2}(1/2^\tau) \sum_{m \in I_\tau} |\hat{f}(m)|^2 \leq C_7(r).$$

Лемма 4. Пусть функция ω_r типа модуля гладкости порядка r удовлетворяет условию $\sum_{\tau=0}^{\infty} \omega_r(1/2^\tau)2^{\tau s/2} < \infty$. Тогда ряд Фурье $\sum_{m \in Z^s} \hat{f}(m)e^{2\pi i(m,x)}$ каждой функции $f \in W_2^{\omega_r}(0, 1)^s$ сходится абсолютно.

Лемма 5[7]. Пусть дано целое число $s \geq 1$. Тогда для каждого целого $N \geq 1$ выполнено следующее утверждение: для любого множества $G \equiv \{m^{(1)}, \dots, m^{(N')}\} \subset Z^s$ такого, что $N' = |G| \geq 2N$ и $|G| \succ_s N$, и для произвольных линейных функционалов l_1, \dots, l_N , определенных, по крайней мере, на множестве всех тригонометрических полиномов со спектром в G найдутся комплексные числа $\{c_k\}_{k=1}^{N'}$, удовлетворяющие условиям $\sum_{k=1}^{N'} |c_k| \geq N$, $\sum_{k=1}^{N'} |c_k|^2 = N$, причем если $\chi(x) = \sum_{k=1}^{N'} c_k e^{2\pi i(m^{(k)}, x)}$, то $l_1(\chi) = 0, \dots, l_N(\chi) = 0$ и $\|\chi\|_{L^\infty} \geq N$, $\|\chi\|_{L^2} = \sqrt{N}$.

Лемма 6. Пусть функция ω_r типа модуля гладкости порядка r удовлетворяет условию $\sum_{\tau=0}^{\infty} \omega_r(1/2^\tau)2^{\tau s/2} < \infty$. Тогда для каждой возрастающей к $+\infty$ последовательности $\{\alpha_N\}_{N \geq 1}$ имеет место равенство

$$\lim_{N \rightarrow \infty} \alpha_N (\bar{\omega}_r^*(1/\alpha_N))^{s/2} = +\infty.$$

3. Основной результат. Исходной в $K(B)\Pi$ – исследовании (см. [4 - 6]) является величина

$$\begin{aligned} \delta_N(\varepsilon_N, D_N, T, F)_Y &= \inf_{(l^{(N)}, \varphi_N) \in D_N} \delta_N(\varepsilon_N, (l^{(N)}, \varphi_N), T, F)_Y \equiv \\ &\equiv \inf_{(l^{(N)}, \varphi_N) \in D_N} \sup_{f \in F, |\gamma_N^{(\tau)}| \leq 1} \left\| (Tf)(\cdot) - \varphi_N(l_N^{(1)}(f) + \gamma_N^{(1)}\varepsilon_N, \dots, l_N^{(N)}(f) + \gamma_N^{(N)}\varepsilon_N; \cdot) \right\|_Y. \end{aligned}$$

Здесь ε_N – неотрицательная последовательность, математическая модель задается посредством оператора $T : F \mapsto Y, X$ и Y – нормированные пространства функций, заданных соответственно на множествах Ω_X и Ω_Y , $F \subset X$ класс функций. Числовая информация

$$l^{(N)} \equiv l^{(N)}(f) = (l_N^{(1)}(f), \dots, l_N^{(N)}(f))$$

объема $N(N = 1, 2, \dots)$ о функции f из класса F снимается с функционалов $l_N^{(1)} : F \mapsto C, \dots, l_N^{(N)} : F \mapsto C$. Алгоритм переработки информации

$$\varphi_N(z_1, \dots, z_N; \cdot) : C^N \times \Omega_Y \mapsto C$$

есть соответствие, которое при всяком фиксированном $(z_1, \dots, z_N) \in C^N$ как функция от (\cdot) есть элемент Y . Пара $(l^{(N)}, \varphi_N)$ определяет вычислительный агрегат восстановления по точной информации о функции $f \in F$, действующий по правилу $\varphi_N(l_N^{(1)}(f), \dots, l_N^{(N)}(f); \cdot)$, а через D_N обозначается подмножество всех пар $(l^{(N)}, \varphi_N)$.

В данной работе при конкретизации

$$F = W_2^{\omega_r}, Tf = f, Y = L^q(0, 1)^s, D_N = L^{(N)} \times \{\varphi_N\},$$

где $2 \leq q \leq \infty, L^{(N)}$ – множество всех векторов $l^{(N)} = \left(l_N^{(1)}, \dots, l_N^{(N)}\right)$, состоящих из N

линейных функционалов $l_N^{(1)} : W_2^{\omega_r} \mapsto C, \dots, l_N^{(N)} : W_2^{\omega_r} \mapsto C$, сформулирована следующая

Теорема. Пусть даны $s \in \mathbb{N}, q \in [2, \infty]$ и функция ω_r типа модуля гладкости порядка r такая, что

$$\sum_{\tau=0}^{\infty} \omega_r(1/2^\tau) 2^{\tau s/2} < \infty, \omega_r(\delta\eta) \leq C_4(r) \omega_r(\delta) \omega_r(\eta)$$

для некоторого $C_4(r) > 0$ и для всех $0 \leq \delta, \eta \leq 1$. Тогда для каждого $N \equiv N(K) = (2K + 1)^s, K = 1, 2, \dots$ справедливы следующие утверждения:

$$\mathbf{K(В)П - 1.} \quad \delta_N(0; L^{(N)} \times \{\varphi_N\}, Tf = f, W_2^{\omega_r})_{L^q} \succ_{s,r,q} \omega_r\left(\frac{1}{N^{1/s}}\right) N^{1/2-1/q};$$

$$\mathbf{K(В)П - 2.} \quad \delta_N(0; L^{(N)} \times \{\varphi_N\}, Tf = f, W_2^{\omega_r})_{L^q} \succ_{s,r,q} \delta_N(\bar{\varepsilon}_N; (\bar{l}^{(N)}, \bar{\varphi}_N), Tf = f, W_2^{\omega_r})_{L^q} \text{ и}$$

$$\overline{\lim}_{N \rightarrow +\infty} \frac{\delta_N(\eta_N \bar{\varepsilon}_N; (\bar{l}^{(N)}, \bar{\varphi}_N), Tf = f, W_2^{\omega_r})_{L^q}}{\delta_N(0; L^{(N)} \times \{\varphi_N\}, Tf = f, W_2^{\omega_r})_{L^q}}$$

для любой сколь угодно медленно возрастающей к $+\infty$ положительной последовательности $\{\eta_{N(K)}\}_{K \geq 1}$, где $\bar{\varepsilon}_N = \frac{1}{\sqrt{N}} \omega_r\left(\frac{1}{N^{1/s}}\right)$, вычислительный агрегат

$(\bar{l}^{(N)}, \bar{\varphi}_N)$ состоит из функционалов $\bar{l}_N^{(\tau)}(f) = \hat{f}(\bar{m}^{(\tau)})$, $\tau = 1, \dots, N$ и функции

$$\bar{\varphi}_N(z_1, \dots, z_N; x) = \sum_{\tau=1}^N z_\tau e^{2\pi i(\bar{m}^{(\tau)}, x)}, \text{ а } s\text{- мерные целочисленные векторы } \bar{m}^{(1)}, \dots, \bar{m}^{(N)}$$

такие, что $\bar{m}^{(i)} \neq \bar{m}^{(j)}$ при $i \neq j$ и $\bigcup_{\tau=1}^N \{\bar{m}^{(\tau)}\} = A_K, A_K = \{m \in Z^s : |m_1| \leq K, \dots, |m_s| <$

$K\}$;

К(В)П - 3. Для всякого вычислительного агрегата

$$(l^{(N)}, \varphi_N)(x) \equiv \varphi_N\left(\hat{f}(m^{(1)}), \dots, \hat{f}(m^{(N)}); x\right), N = N(K)$$

при любой сколь угодно медленно возрастающей к $+\infty$ положительной последовательности $\{\eta_{N(K)}\}_{K \geq 1}$ имеет место равенство

$$\overline{\lim}_{K \rightarrow +\infty} \frac{\delta_N\left(\eta_N \bar{\varepsilon}_N; \varphi_N\left(\hat{f}(m^{(1)}), \dots, \hat{f}(m^{(N)}); x\right), Tf = f, W_2^{\omega_r}\right)_{L^q}}{\delta_N(0; L^{(N)} \times \{\varphi_N\}, Tf = f, W_2^{\omega_r})_{L^q}}.$$

Замечание. Условию $\omega_r(\delta\eta) \leq C_4(r) \omega_r(\delta) \omega_r(\eta)$, кроме степенных функций, удовлетворяют функции

$$\omega_r(\delta) = \begin{cases} \delta^r \ln^{r_1}(M/\delta), & \text{если } \delta \in (0, 1]; \\ 0, & \text{если } \delta = 0 \end{cases}$$

с параметрами $r > 0, 0 < r_1 < +\infty$. Заметим, что эти функции при достаточно больших $M > 0$ являются функциями типа модуля гладкости порядка r (см., напр.[8]).

Список литературы

- 1 Утесов А.Б. О погрешности квадратурных формул на некоторых классах функций // Вестник КазГНУ им. аль – Фараби, серия Математика. Механика. Информатика. -2000. -Т. 22. №3. С. 19-25.
- 2 Утесов А.Б. Задача восстановления функций и интегралов на обобщенных классах и решений уравнения теплопроводности: дис. . . . канд. физ.- мат. наук, Алматы, 2001.
- 3 Абикенова Ш.К., Утесов А.Б., Темиргалиев Н. О дискретизации решений волнового уравнения с начальными условиями из обобщенных классов Соболева // Матем. заметки. -2012. -Т. 91. №3. С. 459-463.
- 4 Темиргалиев Н., Жубанышева А.Ж. Теория приближений, вычислительная математика и численный анализ в новой концепции в свете Компьютерного (вычислительного) перечника // Вестник ЕНУ имени Л.Н.Гумилева. Серия Математика. Информатика. Механика. -2018. -Т.124. №3. -С. 8 – 88.

- 5 Темиргалиев Н., Жубанышева А.Ж. Компьютерный (вычислительный) поперечник в контексте общей теории восстановления// Известия ВУЗов, Математика. -2019. -№1. С. 89–97.
- 6 Жубанышева А.Ж., Темиргалиев Н. Информативная мощность тригонометрических коэффициентов Фурье и их предельная погрешность при дискретизации оператора дифференцирования на многомерных классах Соболева// Ж. вычисл. матем. и матем. физ. -2015. -Т. 55. №9. -С. 1474 – 1985.
- 7 Ажгаллиев Ш. О дискретизации решений уравнения теплопроводности// Матем. заметки. 20007. -Т. 82. №2. -С. 177 – 182.
- 8 Ульянов П.Л. Об абсолютной сходимости тригонометрических рядов Фурье// Докл. РАН. -1992. -Т. 322. №2. -С. 253 – 258.

А.Б. Утесов

Қ. Жұбанов атындағы Ақтөбе өңірлік университеті, Ә. Молдағұлова даңғ., 34, Ақтөбе, 030000, Қазақстан

Жалпыланған Соболев кластары функцияларын қалыптастыру есебінің толық $K(E)D$ -зерттеуі

Аннотация: Бұл жұмыста жалпыланған Соболев $W_2^{\omega r}$ кластары функцияларын қалыптастыру есебінің толық $K(E)D$ -зерттеуі қалыптастырылуға тиіс f функциясынан алынған N көлемдегі сандық мәлімет сызықтық функционалдардың мәндері болатын жағдайда жүргізілген. Дәл айтқанда, біріншіден, $W_2^{\omega r}$ кластары функцияларын $L^q, 2 \leq q \leq \infty$ метрикада жуықтаудағы қателіктің дәл реті анықталған; екіншіден, дәл ретті жүзеге асыратын нақты есептеу агрегаты ұсынылған және оның дәл ретті сақтайтын, реті бойынша жақсармайтын $\bar{\varepsilon}_N$ шектік қателігі табылған; үшіншіден, қалыптастырылуға тиіс функцияның Фурье коэффициенттері бойынша құрылған кез келген есептеу агрегатының шектік қателігі $\bar{\varepsilon}_N$ шектік қателігінен жақсы болмайтынына (реті бойынша) көз жеткізілген.

Түйін сөздер: $K(E)D$ - зерттеуі, сызықтық функционал, есептеу агрегаты, жуықтау қателігінің дәл реті, жалпыланған Соболев класы.

A.B. Utesov

K.Zhubanov Aktobe Regional University, Moldaguova ave., 34, Aktobe, 030000, Kazakhstan

Full $C(N)D$ -research of the problem of recovery functions from the generalized Sobolev class

Abstract: In this paper a complete $C(N)D$ -research of the problem of recovery functions from the generalized Sobolev class $W_2^{\omega r}$ is carried out in the case, where numerical information of volume N about the function f being restored is removed from linear functionals. Namely, firstly, the exact order of error of restoring functions from classes $W_2^{\omega r}$; is established in the metric $L^q, 2 \leq q \leq \infty$; secondly, a specific computing unit is proposed, that implements the exact order and its limiting error $\bar{\varepsilon}_N$ is found, that preserves the exact order and not improved in order; thirdly, it is proved that any computing unit constructed by the Fourier coefficients of the function being restored does not have a limiting error, better (in order) than $\bar{\varepsilon}_N$.

Keywords: $C(N)D$ – research, linear functional, computing unit, exact order of error of restoring, generalized Sobolev class.

References

- 1 Utesov A.B. O pogreshnosti kvadraturnyh formul na nekotoryh klassah funkciy [On the error of quadrature formulas on some classes of functions], Bulletin of al-Farabi Kazakh State National University. Mathematics. Mechanics. Computer Science Series. 2000. Vol. 22. №3. P. 19-25. [in Russian]
- 2 Utesov A.B. The problem of recovering functions and integrals on generalized classes and solutions of the heat equation [Zadacha vosstanovleniya funkciy i integralov na obobshchennyh klassah i reshenij uravneniya teploprovodnosti]: Cand. Sci. (Phys.– Math.) Dissertation, Almaty, 2001.
- 3 Abikenova, S., Utesov, A., Temirgaliev, N. On the discretization of solutions of the wave equation with initial conditions from generalized Sobolev classes, Mathematical Notes. 2012. Vol. 91. Is. 3., P. 430–434. DOI: <https://doi.org/10.1134/S0001434612030121>
- 4 Temirgaliev N., Zhubanisheva A.Zh. Teoriya priblizhenij, vychislitel'naya matematika i chislennyj analiz v novoj koncepcii v svete Komp'yuternogo (vychislitel'nogo) poperechnika [Approximation Theory, Computational Mathematics and Numerical Analysis in newconception of Computational (Numerical) Diameter], Bulletin of L.N. Gumilov Eurasian National University. Mathematics. Computer Science. Mechanics Series. 2018. Vol. 124. №3. P. 8 - 88. [in Russian]
- 5 Temirgaliev N., Zhubanisheva A.Zh. Computational (Numerical) diameter in the context of general theory of a recovery, Russian Mathematics (Iz. VUZ). 2019. №1. P. 89-97. DOI: <https://doi.org/10.26907/0021-3446-2019-1-89-97>
- 6 Temirgaliev N., Zhubanisheva A.Zh. Informative Cardinality of Trigonometric Fourier Coefficients and Their Limiting Error in the Discretization of a Differentiation Operator in Multidimensional Sobolev Classes, Computational Mathematics and Mathematical Physics. 2015. Vol. 55. №9. P. 1432-1443. DOI: <https://doi.org/10.1134/S0965542515090146>
- 7 Azhgalliev Sh. On the discretization of solutions of the heat equation, Mathematical Notes. 2007. Vol. 82. №2. P. 153– 158. DOI: <https://doi.org/10.4213/mzm3789>
- 8 Ulyanov P.L. Ob absolyutnoj skhodimosti trigonometricheskikh ryadov Fur'e [On the absolute convergence of trigonometric Fourier series], Dokl. Math. 1992. Vol. 322. №2. P. 253–258. [in Russian]

Сведения об авторе:

Утесов Адилжан Базарханович – кандидат физико–математических наук, доцент кафедры «Математика» Актобинского регионального университета им. К. Жубанова, ул. А. Молдагуловой, 34, Актобе, 030000, Казахстан.

Utessov Adilzhan – candidate of physico - mathematical sciences, associate professor of the Mathematic department K.Zhubanov Aktobe Regional University, A.Moldagulova ave., 34, Aktobe, 030000, Kazakhstan.

Поступила в редакцию 11.11.2022

МРНТИ: 28.23.17

Ж.М. Кожирбаев

*National Laboratory Astana, пр. Кабанбай батыр, 53, Астана, Казахстан
(E-mail: zhanibek.kozhirbayev@nu.edu.kz)*

Самообучение для улучшения системы распознавания речи на казахском языке¹

Аннотация: В последнее время достижения в нейронных моделях, обученных с использованием обширных многоязычных текстовых и устных данных, продемонстрировали многообещающий потенциал для улучшения ситуации с языками, которым не хватает ресурсов. Это исследование сосредоточено на проведении экспериментов с использованием передовых моделей распознавания речи, в частности, Wav2Vec2.0 и Wav2Vec2-XLSR, применительно к казахскому языку. Основная цель этого исследования — оценить эффективность этих моделей при расшифровке разговорного казахского содержания. Кроме того, исследование направлено на изучение возможности использования данных из других языков для начального обучения и оценку того, может ли уточнение модели с помощью данных целевого языка повысить ее производительность. Таким образом, это исследование предлагает ценную информацию о жизнеспособности использования предварительно обученных многоязычных моделей в контексте языков с ограниченными ресурсами. Точно настроенная модель wav2vec2.0-XLSR показала исключительные результаты, продемонстрировав коэффициент ошибок символов (CER) 1,9 и коэффициент ошибок слов (WER) 8,9 при сравнении с тестовым набором данных kazcorpus. Результаты этого анализа могут способствовать созданию надежных и эффективных систем автоматического распознавания речи (ASR), адаптированных для казахского языка. Эти разработки принесут пользу целому ряду приложений, в том числе преобразованию речи в текст, голосовым помощникам и средствам голосового общения.

Ключевые слова: автоматическое распознавание речи, казахский язык, Wav2Vec 2.0, Wav2Vec2-XLSR, предварительно обученные модели-трансформеры, модели представления речи.

DOI: <https://doi.org/10.32523/2616-7182/bulmathenu.2023/4.2>

2000 Mathematics Subject Classification: 68T10

1. ВВЕДЕНИЕ

В последнее время модели на основе последовательностей продемонстрировали замечательные достижения в распознавании речи при сопоставлении с традиционными системами автоматического распознавания речи. Эти основанные на последовательности модели используют нейронные сети для преобразования речи в текст, что упрощает процесс моделирования. Среди них архитектура глубоких нейронных сетей Трансформеры [21] выделяется как широко распространенный и демонстрирует заслуживающие внимания достижения в построении сквозных систем распознавания речи [2–4]. Несмотря на значительный прогресс в развитии моделей ASR, разработка надежных моделей для языков, помимо английского, остается сложной задачей. Эта проблема в основном возникает из-за того, что современные модели, как правило, требуют длительных часов

¹Работа выполнена при финансовой поддержке Министерства науки и высшего образования Республики Казахстан (проект №AP13068635)

аннотированных речевых данных для обучения, чтобы достичь удовлетворительных результатов. Это особенно верно для казахского, тюркского языка, на котором говорит мировое сообщество, насчитывающее более 13 миллионов человек (согласно статистике веб-сайта Ethnologue²).

Недавние успехи в методологиях обучения с самоконтролем продемонстрировали потенциал в решении проблемы ограниченной доступности данных для языков без достаточных ресурсов. Обучение с самостоятельным наблюдением — уникальный подход к совершенствованию систем распознавания речи — использует множество немаркированных речевых данных для получения ценных представлений речевого сигнала. В отличие от традиционного контролируемого обучения, которое требует помеченных данных для обучения модели, алгоритмы самоконтролируемого обучения извлекают знания из необработанных данных без необходимости явных аннотаций. В рамках обучения с самостоятельным наблюдением модели оттачиваются для выполнения задач, тесно связанных с основной задачей распознавания речи, но лишенных необходимости в помеченных входных данных. Метод включает в себя обучение модели прогнозированию последующего кадра речевого сигнала с учетом предыдущих кадров — концепция, называемая контрастным прогностическим кодированием (CPC) [5].

Другая стратегия влечет за собой обучение модели различению двух отдельных речевых сегментов, например, различие между парой близких во времени речевых кадров и теми, которые значительно разнесены друг от друга. Потенциал обучения с самостоятельным наблюдением в сфере распознавания казахской речи заключается в его способности извлекать выгоду из значительных объемов неразмеченных данных, что является значительным преимуществом для языков, где аннотированные данные остаются дефицитными. Используя методы самоконтроля для обучения модели распознавания речи на немаркированных данных, модель может умело расшифровывать ключевые аспекты речевого сигнала, включая фонемы и акустические нюансы, необходимые для точной транскрипции. Помимо своего потенциала для повышения точности систем распознавания речи, обучение с самоконтролем обещает уменьшить количество необходимых помеченных данных для обучения. Это сокращение может существенно сократить затраты и время, необходимые для создания надежной системы распознавания казахской речи, сделав ее более доступной для исследователей и разработчиков. Недавние усовершенствования в кодировщиках звука с самоконтролем, примером которых является Wav2Vec2.0 [6], эффективно ассимилировали высококласные аудиопредставления. Однако их неконтролируемый подход к предварительному обучению создает проблему эффективного преобразования этих представлений в практические результаты. Следовательно, фаза тонкой настройки становится обязательной для профессионального развертывания этих моделей для таких задач, как автоматическое распознавание речи. Существуют две опции предварительно обученных моделей Wav2Vec2.0: первая предварительно обучена исключительно для одного языка, а вторая — многоязычная предварительно обученная модель (XLSR-53). Это исследование было предпринято с целью противопоставления этих двух подходов, чтобы установить их эффективность в обеспечении надежного ASR для казахского языка. Основные вклады этого исследования включают:

1. В дополнение к имеющимся речевым корпусам на казахском языке мы собрали аудиозаписи вместе с соответствующими им текстами из общедоступных источников. Совокупные данные, собранные примерно за 1000 часов. Каждому аудиофайлу соответствовал отдельный текстовый файл, содержащий содержание аудиокниги. Примечательно, что необходимо признать, что между звуком и текстом отсутствовала синхронизация, что подразумевает отсутствие выравнивания на уровне предложения или слова. Чтобы решить эту проблему рассогласования, мы использовали технику сегментации, основанную на алгоритме коннекционистской временной классификации (CTC) [7]. Этот подход облегчил точное извлечение выравнивания аудио-текста.

²<https://www.ethnologue.com/language/kaz>

2. Была проведена серия экспериментов с использованием базовой архитектуры Wav2Vec2.0 и XLSR-53, охватывающих различные сценарии предварительной подготовки и тонкой настройки.
3. Был проведен обширный сравнительный анализ двух методологий: Wav2Vec2.0 и Wav2Vec2.0-XLSR.

2. ОСНОВНАЯ ЧАСТЬ

В этом разделе представлен краткий обзор соответствующей литературы, связанной с этой статьей, разделенный на два подраздела: Wav2Vec2.0 и XLSR-53, а также Казахский ASR.

Wav2Vec 2.0. тщательно разработан для расшифровки речи, встроенной в аудиосигналы, с использованием самоконтролируемой методологии предварительного обучения, которая впитывает идеи из огромных объемов немаркированных аудиоданных. Он объединяет принципы нескольких предшествующих моделей, а именно, Contrastive Predictive Coding (CPC) [5], Model Predictive Control (MPC) [8], wav2vec [9] и vqwav2vec [10]. Архитектура Wav2Vec2 гармонизирует сверточные нейронные сети (CNN) и преобразователи, что позволяет ему воспринимать как локальные нюансы, так и всеобъемлющие шаблоны в аудиоданных. В модели используется многоуровневый сверточный кодировщик признаков, обозначенный как $f : X \rightarrow Z$, для кодирования необработанных аудиосигналов X в представления скрытой речи z_1, \dots, z_T , которые затем подаются в сеть с масками преобразователя, обозначаемые как $g : Z \rightarrow C$, который отображает представления из скрытого пространства в дискретный набор выходных данных, q_1, \dots, q_T , которые представляют цели в самоконтролируемой цели обучения [6, 12]. Модуль преобразования контекстуализирует квантованные представления с помощью блоков внимания, в результате чего получается набор дискретных контекстуальных представлений c_1, \dots, c_T . Кодировщик функций состоит из семи сверточных блоков, каждый из которых имеет 512 каналов, ширину ядра $\{10, 3, 3, 3, 3, 2, 2\}$ и шаги $\{5, 2, 2, 2, 2, 2, 2\}$. С другой стороны, трансформаторная сеть состоит из 24 блоков с 1024 размерами и 4096 внутренними размерами. Всего в ней также имеется 16 головок внимания.

XLSR-53 [11] представляет собой многоязычную модель, основанную на межъязыковой модели XLM-R, предназначенную для решения задач многоязычной и межъязыковой обработки естественного языка (NLP). Основываясь на модели Wav2Vec 2.0, XLSR-53 обладает способностью получать скрытое квантование, охватывающее различные языки. Это достигается за счет использования квантования произведения для отбора квантованных представлений из кодовых книг. В процессе отбора используется метод Gumbel-Softmax, обеспечивающий полное различие. Архитектура XLSR-53 похожа на архитектуру двунаправленный кодировщик представлений трансформера (BERT) [12] с заметным отличием: она включает в себя 53 языковых вложения для каждого из поддерживаемых языков. Этот сложный дизайн позволяет модели обрабатывать различные языки, улавливая их тонкости даже в случаях схожего написания или произношения. Кроме того, XLSR-53 может похвастаться огромным количеством параметров, составляющим 500 миллионов, что делает его одной из крупнейших доступных многоязычных моделей. Эта модель обучается на обширном и разнообразном корпусе, включающем текстовые данные речи, извлеченные из более чем 53 языков. Присущая XLSR-53 способность понимать несколько языков делает его исключительно выгодным для межъязыкового трансферного обучения. Это влечет за собой адаптацию модели, обученной на одном языке, для хорошей работы на другом языке, требуя лишь минимального дополнительного обучения.

Недавние достижения в области ASR открыли новые сквозные архитектуры, демонстрирующие впечатляющую точность при наличии достаточного количества наборов данных. Основной принцип, лежащий в основе этих сквозных моделей, вращается вокруг прямого преобразования входных речевых сигналов в последовательности символов. Такой оптимизированный подход оптимизирует процедуры обучения,

тонкой настройки и логического вывода. В области исследований ASR эксперты преимущественно тяготеют к двум различным методологиям обучения систем ASR: полностью контролируемым и самоконтролируемым моделям. В контексте первой категории Есенбаев [13] провели обширное исследование, направленное на преодоление проблемы автоматического, независимого от говорящего распознавания слитной казахской речи, сосредоточив внимание на определенной лексической основе в условиях шумной среды. Предложенная авторами система продемонстрировала похвальные результаты в различных задачах, включая фонетическое распознавание английской речи, а также распознавание непрерывной казахской речи. Примечательно, что система продемонстрировала относительное улучшение качества распознавания до 20%. Обращает на себя внимание достижение качества распознавания 94,5% для казахской речи. По сути, это исследование выступает в качестве основополагающего шага, закладывающего основу для последующей разработки более продвинутых платформ, предназначенных для непрерывного распознавания казахской речи. В своем исследовании авторы [14] углубляются в распознавание потоковой речи посредством реализации модели RNN-T. Эта архитектура построена с использованием нейронных сетей, таких как LSTM и BLSTM, с использованием обучающего набора данных продолжительностью более 300 часов, включающего как подготовленные (чтение), так и записи спонтанной речи. Результаты исследования подчеркивают способность модели RNN-T достигать CER 10,6. В отдельном исследовании тех же авторов Мамырбаев [15] представляет гибридную модель Transformer + CTC (Connectionist Temporal Classification). Он был отточен с использованием набора речевых данных, охватывающего 400 часов. Примечательно, что результаты исследования подчеркивают эффективность модели, регистрируя CER 3,7 и WER 8,3. Безусловно, следует отметить совместную работу исследователей Центра речевых технологий Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики совместно с Костанайским государственным университетом им. А. Байтурсынова [16]. Эти исследователи приступили к проекту, сосредоточенному на признании и синтезе двуязычного (казахско-русского) языка. Их усилия были сосредоточены на развитии области обработки двуязычной речи, тем самым способствуя более широкому спектру исследований в области языковых технологий. Авторы [17] внесли значительный вклад в исследование распознавания речи на казахском языке в виде базы данных KSC 1. Эта обширная база данных служит открытым эталоном, содержащим около 332 часов расшифрованного аудио. База данных включает более 153 000 высказываний, произнесенных людьми из разных возрастных групп, регионов и полов. Авторы использовали сквозную модель (E2E) на основе Transformer, достигнув коэффициента ошибок символов (CER) 2,8 и коэффициента ошибок слов (WER) 8,7 в этом наборе данных. В родственной разработке Муссаходжаева [18] расширили базу данных KSC до ошеломляющих 1128 часов. Это расширение включало в себя включение дополнительных данных из различных источников, включая телевизионные новости, теле- и радиопрограммы, парламентские выступления и подкасты. Авторы тщательно определили спецификации корпуса и обосновали его полезность, используя модель ASR на основе Transformer. Эта модель дала многообещающие результаты: общая частота ошибок в словах (WER) составила 15,1 в проверочном наборе и 15,6 в тестовом наборе. Эти усилия значительно расширяют ресурсы, доступные для развития технологии распознавания речи на казахском языке. Хотя продолжающаяся разработка моделей продемонстрировала выдающееся мастерство, значительная часть из них в значительной степени зависит от методологий обучения с учителем, что требует значительных объемов аннотированных данных. К сожалению, процесс маркировки и аннотирования данных является ресурсоемким, дорогостоящим и трудоемким, часто влекущим за собой ручное вмешательство. Более того, могут возникнуть обстоятельства, когда получение таких данных становится нецелесообразным из-за ограничений или отсутствия. В отличие от области полностью контролируемых моделей, недавние исследования были направлены на использование существенных акустических моделей, обученных с помощью методов самоконтролируемого обучения и обширных резервуаров немаркированных

данных. В качестве примера можно привести работу [19], в которой представили неконтролируемое предварительное обучение с использованием Wav2Vec2.0. Авторы интегрировали факторизованный уровень TDNN, чтобы поддерживать связь между голосом и временными шагами, тем самым повышая эффективность распознавания речи для казахского языка. Кроме того, они использовали многоязычную предварительную подготовку и методы синтеза речи для дальнейшего повышения производительности. Результаты их экспериментов подчеркнули преимущества ассимиляции немаркированных данных из языков за пределами целевого и использования улучшения данных посредством синтеза речи. Эти подходы, в частности, привели к существенному снижению частоты ошибок в словах в наборах тестов. Это исследование знаменует собой заметный шаг в направлении оптимизации систем распознавания речи при одновременном снижении зависимости от обширных помеченных наборов данных.

3. МЕТОДОЛОГИЯ

Этот раздел сосредоточен на наборах данных, специально предназначенных для распознавания речи на казахском языке. Он также углубляется в методологии, используемые для разработки модулей распознавания речи, предназначенных для этого языкового контекста.

ISSAI KSC. Набор данных ISSAI KSC является наиболее обширным открытым ресурсом, созданным для поддержки казахстанских приложений обработки речи и языка [17]. Этот существенный набор данных включает более 332 часов контента, собранного через веб-платформу, предназначенную для записи речи. Эта платформа приглашала добровольцев формулировать предложения, взятые из целого ряда источников, включая книги, законы, Википедию, новостные порталы и блоги. Набор данных KSC может похвастаться разнообразием, включая динамики и аудиозаписи из разных регионов Казахстана с использованием различных устройств, таких как смартфоны, планшеты и ноутбуки. Пул спикеров родом из пяти разных регионов, при этом наборы для проверки и тестирования включают 51,7% спикеров-женщин и 48,3% спикеров-мужчин.

Kazcorpus. Акустический корпус kazcorpus [20] включает в себя два отдельных подкорпуса: kazspeechdb и kazmedia. На основе корпуса kazspeechdb была заложена основа для построения корпуса новостей вещания. Этот подкорпус состоит из речевых фрагментов, а именно 12 675 предложений на казахском языке, записанных в контролируемых студийных условиях. Спикеры принадлежат к разным полам, возрастам и регионам Казахстана. Подкорпус охватывает в общей сложности 22 часа речи с участием 169 говорящих, включая 73 мужских и 96 женских голосов. Каждый спикер произнес по 75 предложений.

С другой стороны, подкорпус KazMedia включает в себя аудио- и текстовые данные, взятые с официальных сайтов телеинформационных агентств, в частности «Хабар», «Астана ТВ» и «31 канал». Текстовые данные представляют собой совокупность новостных статей на казахском языке, публикуемых на официальных сайтах этих каналов. Аудиоданные состоят из файлов WAV, которые представляют собой звуковые дорожки, извлеченные из различных сегментов видеонews, транслируемых на этих каналах на казахском языке. В совокупности этот подкорпус включает 21 час речи.

KazLibriSpeech. Мы собрали аудиозаписи в сочетании с соответствующими текстами из открытых источников, в общей сложности около 1000 часов данных. Каждый аудиофайл выравнивается с соответствующим текстом в аудиокниге, хотя выравнивания на уровне предложения или слова нет. Следовательно, последующая попытка состоит в том, чтобы разделить эти аудиофайлы на более мелкие интервалы, будь то слова, фразы или предложения. Затем каждый такой сегмент необходимо сравнить с соответствующим ему озвученным текстом, воспроизведенным в том же интервале. Хотя процесс выравнивания и сегментации может быть сложным, этот метод позволяет создавать существенные наборы данных, охватывающие различные источники и домены, с минимальными затратами.

Учитывая, что качество собранных аудиокниг варьируется, был начат процесс очистки и нормализации. Это включало удаление шума, управление омоглифами, транслитерацию, извлечение невоспроизводимых фрагментов, замену акронимов и аббревиатур их полными формами, числовую нормализацию, замену символов их фонетическими аналогами, первоначальную сегментацию на уровне глав и разделение текста. Преобразование исходного текста в краткие предложения с использованием знаков препинания. Кроме того, любые сопутствующие музыкальные элементы в начале и конце аудиокниг были удалены.

Наш подход к сегментации заключался в использовании алгоритма CTC, который обеспечивает точное выравнивание аудиотекста, даже если аудиозапись включает неразборчивые речевые фрагменты в начале или в конце. Наш метод включает в себя обучение сквозной сети на предварительно выровненных данных с использованием системы ASR на основе CTC-внимания. CTC, как механизм вывода и оценки нейронной сети, играет важную роль в обучении рекуррентных нейронных сетей для решения задач, основанных на последовательности, с учетом переменного времени. Этот механизм не зависит от базовой структуры нейронной сети. В нашем контексте эта модель разграничивает речевые сегменты в аудиофайлах на уровне предложений. Модель ASR, необходимая для сегментации, была отточена с использованием набора данных ISSAI KSC в инструменте Espnet [21].

Более подробное описание имеющихся корпусов, предназначенных для распознавания речи на казахском языке, приведено в Таблице 1.

Таблица 1 – Структура корпусов казахского языка

Структура	Наименование корпуса/наборов	Тип данных	Количество wav-файлов	Общая продолжительность wav-файлов (час)
1	ISSAI KSC		153853	332.6
1.1	Train	Краудсорсинговые записи	147236	318.4
1.2	Dev		3283	7.1
1.3	Test		3334	7.1
2	Kazcorpus	Смешанный тип: студийные записи, подготовленная речь + спонтанная речь в разных акустических условиях	13425	44.16
2.1	kazspeechdb		12675	22.61
2.1.1	Train		11175	19.92
2.1.2	Dev		750	1.36
2.1.3	Test		750	1.34
2.2	KazMedia		740	21.55
2.2.1	Train		561	18.04
2.2.2	Dev		49	1.00
2.2.3	Test		130	2.51
3	KazLibriSpeech	Аудиокниги	575243	992

В контексте данного исследования была проведена оценка двух вариантов wav2vec: (1) Wav2Vec2.0, предварительно обученный и настроенный исключительно для казахского языка, и (2) XLSR-53, изначально предварительно обученный для 53 языка с последующим непрерывным предварительным обучением и доводкой на казахский язык.

Wav2Vec 2.0. Эксперимент проводился с использованием платформы Fairseq. Базовая модель Wav2Vec 2.0 прошла предварительное обучение с использованием различных конфигураций, включающих в себя уровень отбрасывания кодировщика, установленный на 0,05, dropout_input, dropout_features и feature_grad_mult, установленный на 0,1, и encoder_embed_dim, установленный на 768. Гиперпараметры обучения включали скорость обучения $5 * 10^{-4}$, с фазой разминки в первые 10% продолжительности тренировки. Количество обновлений было указано как 800 000, а максимальное количество токенов – 1 200 000. Последовательно применялся оптимизатор Адама в соответствии с исходным подходом.

Для тонкой настройки использовались стандартные процедуры с параметрами, заданными следующим образом: количество обновлений достигло 160 000, а максимальное

количество токенов составило 2 800 000. Как и в случае с предварительным обучением, был задействован оптимизатор Adam, использующий скорость обучения $3 \cdot 10^{-5}$ и накопление градиента из 12 шагов. Размер пакета обучения динамически определялся платформой с учетом заданного максимума маркеров. В ходе обучения оптимальная модель выбиралась на основе наименьшего значения WER, достигнутого в проверочном наборе.

XLSR-53. Модель XLSR прошла предварительную подготовку с идентичными конфигурациями, используемыми для большой модели Wav2Vec. Блок кодера состоял из 24 слоев, каждый размером 1024, и использовался набор из 16 блоков внимания без учета исключения. Параметры тонкой настройки были определены в соответствии с конфигурациями, примененными в исходном эксперименте XLSR с Wav2Vec 2.0.

Языковая модель. После процесса тонкой настройки модель подвергается декодированию с помощью 3-граммовой языковой модели. Языковая модель была обучена с использованием транскрипций из всех доступных наборов данных, указанных в таблице 1, с использованием набора инструментов Kenlm. Для целей декодирования используется декодер поиска луча с размером луча, настроенным на 1500.

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

При оценке моделей Wav2Vec2.0 использовались наборы данных, описанные в разделе 3. Производительность систем ASR, указанная в баллах WER и CER, представлена в таблице 2. Для каждой архитектуры использовались различные сценарии обучения с разными параметрами. Эксперименты проводились на сервере NVIDIA DGX-1, оснащенный 8 графическими процессорами V100.

Таблица 2 – Производительность моделей Wav2Vec 2.0

ID	Исходная модель	Набор данных предварительной подготовки	Набор данных тонкой настройки	Оценочный набор	Набор ЯМ	Тест WER	Тест CER
1	Wav2Vec 2.0 Base	KazLibriSpeech	Kazcorpus (train+dev)	Kazcorpus (test)	3-gram KenLM	12.4	2.6
2	Wav2Vec 2.0 Base	KazLibriSpeech	ISSAI KSC1 (train+dev)	ISSAI KSC1 (test)	3-gram KenLM	10.1	2.8
3	XLSR-53	KazLibriSpeech	Kazcorpus (train+dev)	Kazcorpus (test)	3-gram KenLM	8.9	1.9
4	XLSR-53	KazLibriSpeech	ISSAI KSC1 (train+dev)	ISSAI KSC1 (test)	3-gram KenLM	15.1	4.8

В Таблице 2 показаны баллы частоты ошибок символов и ошибок слов для точно настроенных моделей Wav2Vec 2.0-base и XLSR-53. На этапе предварительного обучения использовался исключительно корпус KazLibriSpeech, а для тонкой настройки использовались ISSAI KSC1 и Kazcorpus. Результаты подчеркивают исключительную производительность предварительно обученной модели XLSR-53. После предварительной подготовки с использованием корпуса KazLi-briSpeech и тонкой настройки с использованием данных Kazcorpus (train+dev) он достигает CER 1,9 и WER 8,9 на тестовом наборе. Эти цифры заметно выше на 28,2% и 26,9% по сравнению с базовой моделью Wav2Vec 2.0 с точки зрения WER и CER соответственно. Эти результаты подчеркивают значительное улучшение производительности модели благодаря предварительному обучению, где размер набора данных, используемого для предварительного обучения, играет ключевую роль. И наоборот, базовая модель Wav2Vec 2.0, прошедшая предварительное обучение с помощью корпуса KazLibriSpeech и тонкую

настройку с использованием данных ISSAI KSC1 (train+dev), демонстрирует превосходные результаты по сравнению с моделью XLSR-53. Это событие могло быть вызвано тем, что исходная модель прошла предварительное обучение с использованием набора данных, содержащего аудиокниги. Ход обучения на этапах предварительного обучения и тонкой настройки показан на Рисунках 1 и 2. Эти визуальные эффекты отображают изменения значений потерь в ходе обучения, предлагая представление о процессе оптимизации моделей.

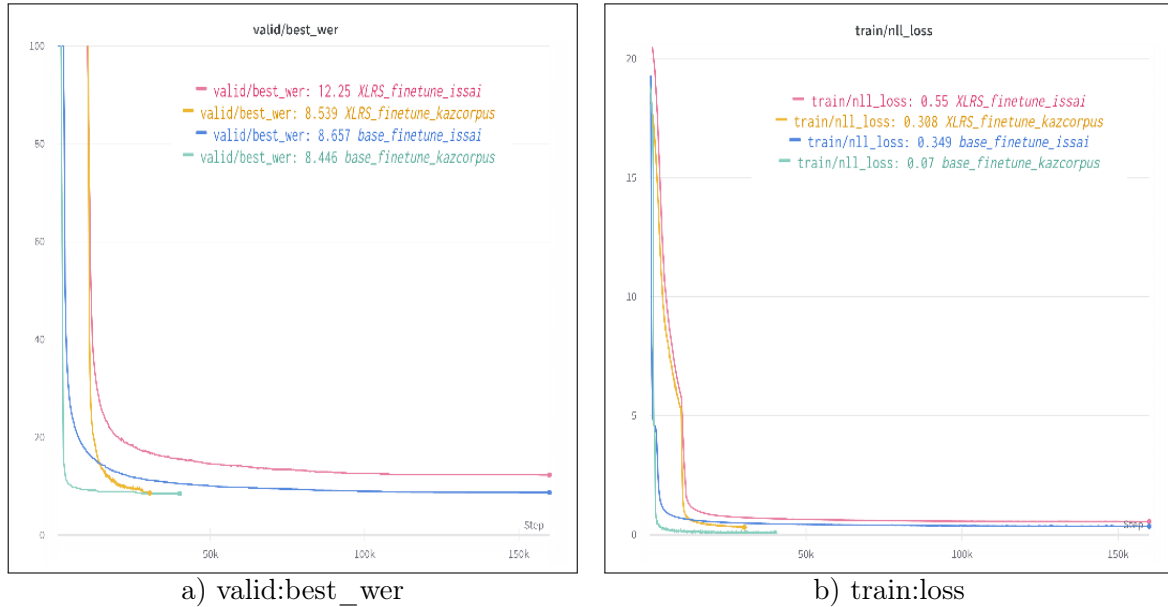


Рисунок 1 – Точная настройка Wav2Vec 2.0

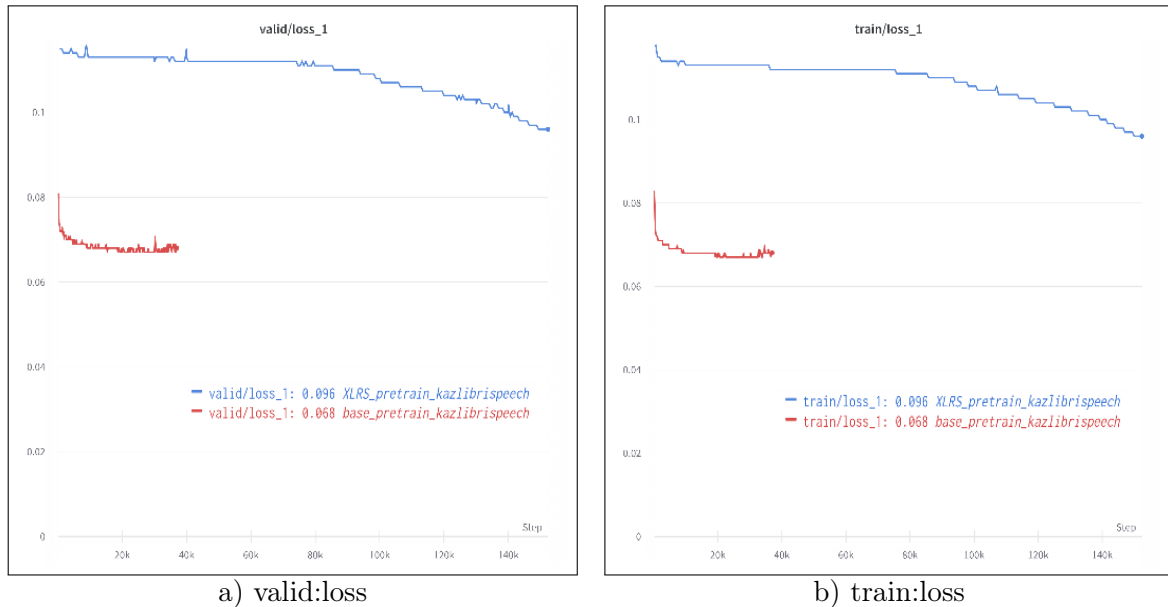


Рисунок 2 – Предварительное обучение Wav2Vec 2.0

5. ВЫВОДЫ

Основная цель этого исследования — оценить эффективность передовых моделей распознавания речи при расшифровке казахского языка, который относится к категории

малоресурсных языков. Наш анализ включает в себя всестороннее сравнение этих моделей с учетом характера и объема данных, используемых на этапах предварительной подготовки и тонкой настройки. Кроме того, исследование направлено на раскрытие потенциальных преимуществ первоначального предварительного обучения с использованием данных из других языков с последующей тонкой настройкой данных из целевого языка. Исследование включает в себя серию экспериментов с использованием архитектур Wav2Vec2.0 и Wav2Vec2-XLS-R, в которых исследуются различные сценарии предварительной подготовки и тонкой настройки. С помощью этих экспериментов мы не только получаем представление о производительности этих моделей специально для казахского языка, но и выясняем последствия, применимые к другим языкам и условиям.

По сути, это исследование освещает перспективы использования усовершенствованных многоязычных моделей и сопоставления самоконтролируемых и полностью контролируемых методов для надежного автоматического распознавания речи в языковых контекстах с ограниченными ресурсами. Выводы и методологии, представленные в этом исследовании, имеют более широкое значение, распространяющееся на смягчение ограничений языковых ресурсов и продвижение разработки системы ASR для широкого круга языков.

Список литературы

- 1 Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. Attention is all you need // Advances in neural information processing systems. - 2017. - V. 30. - P. 6000–6010.
- 2 Karita S., Chen N., Hayashi T., Hori T., Inaguma H., Jiang Z., Someki M., Soplein N., Yamamoto R., Wang X., Watanabe S. A comparative study on transformer vs rnn in speech applications // Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). - Singapore, 2019. - P. 449-456.
- 3 Nakatani T. Improving transformer-based end-to-end speech recognition with connectionist temporal classification and language model integration // Proceedings of Interspeech. - Graz, Austria, 2019. - P. 1408-1412.
- 4 Dong L., Xu S., Xu B. Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition // Proceedings of IEEE international conference on acoustics, speech and signal processing (ICASSP). - Calgary, Canada, 2018. - P. 5884-5888.
- 5 Song J., Ermon S. Multi-label contrastive predictive coding // Advances in Neural Information Processing Systems. - 2020. - V. 33. - P. 8161-8173.
- 6 Baevski A., Zhou Y., Mohamed A., Auli M. wav2vec 2.0: A framework for self-supervised learning of speech representations // Advances in neural information processing systems. - 2020. - V. 33. - P. 12449-12460.
- 7 Kurzinger L., Winkelbauer D., Li L., Watzel T., Rigoll G. CTC-segmentation of large corpora for german end-to-end speech recognition // Proceedings of Speech and Computer: 22nd International Conference (SPECOM). - St. Petersburg, Russia, 2020. - P. 267-278.
- 8 Li S., Li L., Hong Q., Liu L. Improving Transformer-Based Speech Recognition with Unsupervised Pre-Training and Multi-Task Semantic Knowledge Learning // Proceedings of Interspeech. - Shanghai, China, 2020. - P. 5006-5010.
- 9 Schneider S., Baevski A., Collobert R., Auli, M. wav2vec: Unsupervised Pre-Training for Speech Recognition // Proceedings of Interspeech. - Graz, Austria, 2019. - P. 3465-3469.
- 10 Baevski A., Schneider S., Auli M. vq-wav2vec: Self-supervised learning of discrete speech representations // Proceedings of 8th International Conference on Learning Representations (ICLR). - Addis Ababa, Ethiopia, 2020. - P. 1-12.
- 11 Conneau, A., Baevski, A., Collobert, R., Mohamed, A., Auli, M. Unsupervised cross-lingual representation learning for speech recognition // Proceedings of Interspeech. - Brno, Czechia, 2021. - P. 2426-2430.
- 12 Devlin J., Chang M. W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. - Minneapolis, Minnesota, USA, 2019. - P. 4171-4186.
- 13 Yessenbayev Z., Karabalayeva M., Shamayeva F. Large Vocabulary Continuous Speech Recognition for Kazakh // Proceedings of the I International Conference on Computer processing of Turkic Languages. - Astana, Kazakhstan, 2013. - P. 217-221.
- 14 Мамырбайев О., Оралбекова Д., Кыдырбекова А., Турдалыкызы Т., Бекарыстанкызы А. End-to-end model based on RNN-T for Kazakh speech recognition // Proceedings of the 3rd International Conference on Computer Communication and the Internet (ICCCI). - Nagoya, Japan, 2021. - P. 163-167.
- 15 Мамырбайев О., Оралбекова Д., Алимхан К., Нуранбайева В. Hybrid end-to-end model for Kazakh speech recognition // International Journal of Speech Technology. - 2022. - P. 1-10.

- 16 Khomitsevich O., Mendelev V., Tomashenko N., Rybin S., Medennikov I., Kudubayeva S. A bilingual Kazakh-Russian system for automatic speech recognition and synthesis // Proceedings of Speech and Computer: 17th International Conference (SPECOM). - Athens, Greece, 2015. - P. 25-33.
- 17 K7assanov Y., Mussakhoyayeva S., Mirzakhmetov A., Adiyev A., Nurpeiissov M., Varol H. A Crowdsourced Open-Source Kazakh Speech Corpus and Initial Speech Recognition Baseline // Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. - Virtual, 2021. - P. 697-706.
- 18 Mussakhoyayeva S., Khassanov Y., Varol H. KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus // Proceedings of Interspeech. - Incheon, Republic of Korea, 2015. - P. 18-22.
- 19 Meng W., Yolwas N. A Study of Speech Recognition for Kazakh Based on Unsupervised Pre-Training // Sensors. - 2023. - Vol. 23, №2. - P. 870-883.
- 20 Makhambetov O., Makazhanov A., Yessenbayev Z., Matkarimov B., Sabyrgaliyev I., Sharafudinov A. Assembling the kazakh language corpus // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. - Seattle, Washington, USA, 2013. - P. 1022-103.
- 21 Watanabe S., Hori T., Karita S., Hayashi T., Nishitoba J., Unno Y., Soplín N., Heymann J., Wiesner M., Chen N., Renduchintala A. Espnet: End-to-end speech processing toolkit // Proceedings of Interspeech. - Hyderabad, India, 2018. - P. 2207-2211.

Ж.М. Кожирбаев

National Laboratory Astana, Қабанбай батыр даңғылы, 53, Астана, Қазақстан

Қазақ тілін тану жүйесін өз бетінше оқыту

Аннотация: Жақында ауқымды көптілді мәтіндік және ауызша деректерге үйретілген нейрондық модельдердегі жетістіктер ресурсы аз тілдердің жағдайын жақсартудың перспективалы әлеуетін көрсетті. Бұл зерттеу сөзді танудың жетілдірілген үлгілерін, атап айтқанда Wav2Vec2.0 және Wav2Vec2-XLSR, қазақ тіліне қолданылатын эксперименттерді жүргізуге бағытталған. Бұл зерттеудің негізгі мақсаты – осы үлгілердің ауызша қазақша мазмұнды транскрипциялаудағы тиімділігін бағалау. Сонымен қатар, зерттеу бастапқы оқыту үшін басқа тілдердегі деректерді пайдалану мүмкіндігін зерттеуге және үлгіні мақсатты тілдегі деректермен нақтылау оның өнімділігін жақсартуға болатынын бағалауға бағытталған. Осылайша, бұл зерттеу ресурс шектеулі тілдер контекстінде алдын ала дайындалған көптілді модельдерді пайдаланудың өміршеңдігі туралы құнды ақпаратты ұсынады. Жақсы бапталған wav2vec2.0-XLSR моделі kazcorpus деректер жинағының сынақ жинағымен салыстырғанда 1,9 таңба қатесінің көрсеткісін (CER) және 8,9 сөз қатесінің көрсеткісін (WER) бере отырып, өте жақсы жұмыс жасады. Бұл талдаудың нәтижелері қазақ тіліне бейімделген сенімді және тиімді сөзді автоматты түрде тану (ASR) жүйелерін құруға ықпал ете алады. Бұл әзірлемелер сөйлеуден мәтінге, дауыстық көмекшілер мен дауыстық байланыстарды қоса алғанда, бірқатар қолданбаларға пайдалы болады.

Түйін сөздер: автоматты түрде сөйлеуді тану, қазақ тілі, Wav2Vec 2.0, Wav2Vec2-XLSR, алдын ала дайындалған трансформер үлгілері, сөйлеуді бейнелеу үлгілері.

Zh.M. Kozhirbayev

National Laboratory Astana, Kabanbay batyr ave. 53, Astana, Kazakhstan

Self-Supervised Training for the Kazakh Speech Recognition System

Abstract: In recent times, advancements in neural models trained using extensive multilingual textual and spoken data have displayed promising potential for enhancing the situation of languages that lack resources. This study is centered on conducting experiments utilizing cutting-edge speech recognition models, specifically Wav2Vec2.0 and Wav2Vec2-XLSR, applied to the Kazakh language. The primary aim of this research is to assess the efficacy of these models in transcribing spoken Kazakh content. Additionally, the investigation seeks to explore the feasibility of leveraging data from other languages for initial training, and to assess whether refining the model with target language data can enhance its performance. As such, this study offers valuable insights into the viability of employing pre-trained multilingual models in the context of under-resourced languages. The fine-tuned wav2vec2.0-XLSR model achieved exceptional results, boasting a character error rate (CER) of 1.9 and a word error rate (WER) of 8.9 when evaluated against the test set of the kazcorpus dataset. The outcomes of this analysis hold potential to advance the creation of robust and efficient Automatic Speech Recognition (ASR) systems tailored for the Kazakh language. These developments stand to benefit a range of applications, including speech-to-text translation, voice-activated assistants, and speech-driven communication tools.

Keywords: automatic speech recognition, Kazakh language, Wav2Vec 2.0, Wav2Vec2-XLSR, pre-trained transformer models, speech representation models.

References

- 1 Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. Attention is all you need [Advances in neural information processing systems]. 2017. Vol. 30. P. 6000–6010.
- 2 Karita S., Chen N., Hayashi T., Hori T., Inaguma H., Jiang Z., Someki M., Soplín N., Yamamoto R., Wang X., Watanabe S. A comparative study on transformer vs rnn in speech applications [Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)], Singapore, 2019. P. 449-456.

- 3 Nakatani T. Improving transformer-based end-to-end speech recognition with connectionist temporal classification and language model integration [Proceedings of Interspeech], Graz, Austria, 2019. P. 1408-1412.
- 4 Dong L., Xu S., Xu B. Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition [Proceedings of IEEE international conference on acoustics, speech and signal processing (ICASSP)], Calgary, Canada, 2018. P. 5884-5888.
- 5 Song J., Ermon S. Multi-label contrastive predictive coding [Advances in Neural Information Processing Systems]. 2020. Vol. 33. P. 8161-8173.
- 6 Baevski A., Zhou Y., Mohamed A., Auli M. wav2vec 2.0: A framework for self-supervised learning of speech representations [Advances in neural information processing systems]. 2020. Vol. 33. P. 12449-12460.
- 7 Kurzinger L., Winkelbauer D., Li L., Watzel T., Rigoll G. CTC-segmentation of large corpora for german end-to-end speech recognition [Proceedings of Speech and Computer: 22nd International Conference (SPECOM)], St. Petersburg, Russia, 2020. P. 267-278.
- 8 Li S., Li L., Hong Q., Liu L. Improving Transformer-Based Speech Recognition with Unsupervised Pre-Training and Multi-Task Semantic Knowledge Learning [Proceedings of Interspeech], Shanghai, China, 2020. P. 5006-5010.
- 9 Schneider S., Baevski A., Collobert R., Auli, M. wav2vec: Unsupervised Pre-Training for Speech Recognition [Proceedings of Interspeech], Graz, Austria, 2019. P. 3465-3469.
- 10 Baevski A., Schneider S., Auli M. vq-wav2vec: Self-supervised learning of discrete speech representations [Proceedings of 8th International Conference on Learning Representations (ICLR)], Addis Ababa, Ethiopia, 2020. P. 1-12.
- 11 Conneau, A., Baevski, A., Collobert, R., Mohamed, A., Auli, M. Unsupervised cross-lingual representation learning for speech recognition [Proceedings of Interspeech], Brno, Czechia, 2021. P. 2426-2430.
- 12 Devlin J., Chang M. W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies], Minneapolis, Minnesota, USA, 2019. P. 4171-4186.
- 13 Yessenbayev Z., Karabalayeva M., Shamayeva F. Large Vocabulary Continuous Speech Recognition for Kazakh [Proceedings of the I International Conference on Computer processing of Turkic Languages], Astana, Kazakhstan, 2013. P. 217-221.
- 14 Mamyrbayev O., Oralbekova D., Kydyrbekova A., Turdalykyzy T., Bekarystankyzy A. End-to-end model based on RNN-T for Kazakh speech recognition [Proceedings of the 3rd International Conference on Computer Communication and the Internet (ICCCI)], Nagoya, Japan, 2021. P. 163-167.
- 15 Mamyrbayev O., Oralbekova D., Alimhan K., Nuranbayeva B. Hybrid end-to-end model for Kazakh speech recognition [International Journal of Speech Technology]. 2022. P. 1-10.
- 16 Khomitsevich O., Mendeleev V., Tomashenko N., Rybin S., Medennikov I., Kudubayeva S. A bilingual Kazakh-Russian system for automatic speech recognition and synthesis [Proceedings of Speech and Computer: 17th International Conference (SPECOM)], Athens, Greece, 2015. P. 25-33.
- 17 Khassanov Y., Mussakhoyayeva S., Mirzakhmetov A., Adiyev A., Nurpeissov M., Varol H. A Crowdsourced Open-Source Kazakh Speech Corpus and Initial Speech Recognition Baseline [Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume], Virtual, 2021. P. 697-706.
- 18 Mussakhoyayeva S., Khassanov Y., Varol H. KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus [Proceedings of Interspeech], Incheon, Republic of Korea, 2015. P. 18-22.
- 19 Meng W., Yolwas N. A Study of Speech Recognition for Kazakh Based on Unsupervised Pre-Training [Sensors]. 2023. Vol. 23. №2, P. 870-883.
- 20 Makhambetov O., Makazhanov A., Yessenbayev Z., Matkarimov B., Sabyrgaliyev I., Sharafudinov A. Assembling the kazakh language corpus [Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing], Seattle, Washington, USA, 2013. P. 1022-103.
- 21 Watanabe S., Hori T., Karita S., Hayashi T., Nishitoba J., Unno Y., Soplín N., Heymann J., Wiesner M., Chen N., Renduchintala A. Espnet: End-to-end speech processing toolkit [Proceedings of Interspeech], Hyderabad, India, 2018. P. 2207-2211.

Сведения об авторах:

Кожирбаев Ж.М. – PhD, Старший научный сотрудник, National Laboratory Astana, пр. Кабанбай батыр, 53, Астана, Казахстан.

Kozhirbayev Zh. M. – PhD, Senior Researcher, National Laboratory Astana, Kabanbay batyr ave. 53, Astana, Kazakhstan.

Поступила в редакцию 13.09.2023

IRSTI: 27.43.51

Sh. Abikenova¹, R. Marcelloni², Sh. Aitimova³

¹ National Confederation of Employers "PARYZ", al-Farabi ave., 77/8, Almaty, Kazakhstan

² National Institute for Insurance against Accidents at Work (INAAIL), P. le Pastore, 6, Roma, Italy

³ Republican Research Institute of Labor Protection of the Ministry of Labor and Social Protection of the Population of the Republic of Kazakhstan, Kravtsova str., 18, Astana, Kazakhstan

(E-mail: shabik_29@mail.ru, rafmarce@libero.it, aitimova_80@mail.ru)

Mathematical aspects of occupational risk and its classification considering statistical indicators

Abstract: Occupational risk emerges as a comparatively novel concept within the regulation of social and labor relations in the Republic of Kazakhstan, serving as an indicator of the risk for loss of work capacity or the death of an employee during the execution of their work-related (official) duties due to an occupational accident. This article elucidates the statistical aspects of occupational risk content, affirming the mathematical ubiquity and the validity of the scientific framework employed. The methodological foundation for assessing occupational risk incorporates universally accepted statistical methods for comparison, grouping, and systematisation through the juxtaposition of credible statistical data.

The purpose of the study is to employ one of the mathematical approaches to categorise types of economic activity into classes of occupational risk within the mandatory insurance system against accidents for workers fulfilling their occupational duties. The application of machine learning theory with big data across 132 types of economic activities (two-digit, including some five-digit codes) facilitated the execution of a classification procedure, resulting in the segmentation into 22 classes of occupational risk. It revealed the necessity for introducing a five-digit classification and further detailing the types economic activities as the class of occupational risk increases, such as including the five-digit codes 07101 "Underground mining of iron ores" and 43991 "Mine construction" into class 22 (the highest), delineating from the two-digit codes of the mining and construction industry accordingly. The scientific results were obtained within the framework of program-targeted funding by the Ministry of Labor and Social Protection of the Population of the Republic of Kazakhstan (scientific and technical program No. BR11965728 "Economic problems of safe work and institutional transformations of the insurance mechanism in the Republic of Kazakhstan").

Keywords: occupational risk, labor statistics, frequency of accidents, level of industrial injury, risk relevance sorting (ranks), variability range.

DOI: <https://doi.org/10.32523/2616-7182/bulmathenu.2023/4.3>

2000 Mathematics Subject Classification: 62P05

1. INTRODUCTION

In accordance with the Labor Code of 23 November 2015 No. 414-V ZRC [1], the employer is obliged to insure the employee against accidents in the performance of his/her labor (service) duties. This type of insurance is based on the attribution of types of economic activities (TEA) to classes of occupational risk according to the Regulations of attribution of types of economic

activities to classes of occupational risk approved by the Minister of Labor and Social Protection of the Republic of Kazakhstan dated 5 March 2022 No. 86 [2].

In connection with the introduction in December 2007 of a new classification of TEA (OKED NK RK 03-2019 five-digit), on the basis of the Order of the Committee for Technical Regulation and Metrology of the Ministry of Industry and New Technologies of Republic of Kazakhstan В.,– 683-od from 14.12.2007 [3] in the application of the Regulations of attribution of types of economic activities to classes of occupational risk there were difficulties. It should be recalled that these rules were approved in 2005 as part of the introduction of a new type of insurance - compulsory insurance of workers civil liability for harm caused to the life and health of an workers while performing their work (official) duties [4]. Incidentally, the name of the insurance type was changed in 2010 to compulsory insurance against accidents for workers in the performance of their work (official) duties. While the name of the insurance changed, the model essentially remained unchanged. A comparative analysis of TEA established that the existing classification contains verbal descriptions without specifying numerical codes, and that out of 996 codes in the new classification, only 97 (i.e., only 10%) have verbal descriptions that match the formulations contained in the Regulations for classifying TEA. A significant number of activity types in the new classification are not listed in the rules, and conversely, several verbal descriptions contained in these rules are absent in the new classification.

A few existing codes allow for multiple interpretations, and accordingly, the same OKED (Classification of Economic Activities) can be assigned to different risk classes. This has created a foundation for manipulating risk classes (the possibility of assigning an OKED to the class with the lowest risk and, consequently, with a lower insurance premium among the options), with the aim of attracting a larger number of enterprises. The division of types of activities into risk classes also raises questions when less significant types of activities (for example, the production of ropes, cords, twine, and nets) are given a separate line, while more significant types of activities are "neglected". In this context, the revision of the "Regulations for classifying types of economic activities into Occupational Risk Classes" is a priority task, as is discussed in this article. Scientific results were obtained in accordance with OKED with its verbal description and the indication of the numerical code.

2. METHODS

The materials for the study were based on the results of statistical research on industrial injuries in the Republic of Kazakhstan, conducted during the period from 2018 to 2022, as well as materials from an analytical note on the study of reporting on the number of workers in harmful and other adverse working conditions by specific TEA [5-6].

As a source of statistical information, numerical data collected through Form 7-TPZ "Report on Injuries Related to Occupational Activities and Occupational Diseases" and Form 1-T (working conditions) "Report on the Number of Workers in Harmful and Other Adverse Working Conditions" from the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan (BNS) were used.

Types of economic activity are classified according to the general classifier of economic activities (OKED NK RK 03-2019 five-digit), which covers the complete nomenclature of 996 types of economic activities according to ISIC-Rev.4, developed by the United Nations for the collection, processing, and storage of information necessary for economic analysis and decision-making at the macroeconomic level [7]. In this study, during the analysis of statistical data, both two-digit and five-digit levels of the OKED were utilized. The methodological basis for the research conducted here was guided by the manual [8-9].

3. RESULTS AND DISCUSSIONS

In accordance with the Regulations for Occupational Risk Management, approved by the order of the Minister of Labor and Social Protection of the Population of the Republic of Kazakhstan dated 11 September 2020, No. 363 [10], occupational risk is understood as the risk of loss of work capacity or death of an employee in the performance of work (official) duties. In turn, the Labor Code of the Republic [1] provides that temporary or permanent loss of work capacity or death is caused by an industrial injury, a sudden deterioration in health, or poisoning of the

worker, which are the result of exposure to harmful and/or dangerous production factors while performing work (official) duties, i.e., an accident related to labor activities (at the workplace).

Before proceeding to the results of the scientific research, it is also worth recalling that, according to the legislation of the Republic of Kazakhstan, the concept of an occupational risk class is applied - the level of industrial injuries and occupational diseases that has developed according to TEA [4].

As initial data for classifying TEA into occupational risk classes according to the Regulations for Classifying Types of Economic Activities into Occupational Risk Classes [2], the following statistical indicators over the previous five years are used:

- *average annual listed number of workers by type of economic activity.*
- *number of workers employed in harmful and other adverse working conditions by specific TEA.*
- *number of individuals who have suffered accidents related to work activities.*
- *number of fatalities resulting from work-related accidents.*
- *annual accrued number of payments to victims of work-related accidents, including the average amount of compensation per victim of an accident.*
- *annual total of accrued payments in the event of death of workers related to work activities, including the average size of the accrued payments.*
- *annual total of premiums paid to the insurer for accepting the obligation to make an insurance payout to the beneficiary in the amount determined by the compulsory insurance contract against accidents for the employee.*
- *annual wage fund by type of economic activity.*
- *loss of working time - the number of calendar man-days of incapacity that ended in the reporting year among the victims, including the deceased.*

TEA have been classified into occupational risk categories based on the values of the indicators listed below:

- *proportion of workers employed in harmful and other adverse working conditions by TEA, %.*
- *frequency rate of occupational injuries per 1,000 workers.*
- *frequency rate of fatal accidents per 1,000 workers.*
- *ratio of the frequency rate of fatal accidents per 1,000 workers to the frequency rate of all occupational injuries per 1,000 workers.*
- *distribution of the number of victims by TEA, %.*
- *distribution of the number of workers employed in harmful and other adverse working conditions by TEA, %.*

Also, indicators such as the ratio of the annual accrued amount of payments to victims of work-related accidents and the annual amount of accrued payments in the event of the death of workers related to work activities to the annual wage fund by type of economic activity (%), and the ratio of the annual accrued amount of payments to victims of work-related accidents and the annual amount of accrued payments in the event of the death of workers related to work activities to the annual amount of premiums paid to the insurer by type of economic activity (%).

Given the above, the article examines the issues of assessing the risk of accidents related to work activities. Simultaneously, the application of mathematical methods in assessing the risk of accidents is explored in various articles [11-17].

For instance, Jinxian Weng, Xiafan Gan, Zheyu Zhang proposed a mathematical model, QRA (Quantitative Risk Assessment), for assessing the risk of accidents (in the transportation of hazardous materials at work) [11], which combines the frequency of accidents and their consequences. In the work of Iranian scientists Maryam Sadat Hajakbari and Behrouz Minaei-Bidgoli, the results of a comprehensive thematic study utilizing data mining methods based on data from the Ministry of Labor of Iran are presented, aiming at developing a new scoring system for assessing the risk of workplace accidents [12]. The findings of the research indicate that by integrating methods based on the frequency and severity of accidents, used for assessing organizational safety, and data mining models, the proposed accident risk assessment method

enables the identification and evaluation of critical points with a high level of accuracy. The practical application of the method is demonstrated, identifying the most hazardous workplaces in 2010 and determining the risk level for each category. Subsequently, by repeating the procedure for 2011, critical points are identified based on previous data, and a list of workplaces requiring periodic inspections in 2012 is compiled.

The study [13] is aimed at identifying the most suitable model for predicting the frequency of accidents in the workplace and determining the economic criteria that have the greatest impact. It examines four machine learning algorithms - Random Forest (RF), Support Vector Machine (SVM), Multivariate Adaptive Regression Spline (MARS), and M5 Tree Model (M5), for predicting workplace accidents, considering three economic criteria: Basic Income (BI), Inflation Index (II), and Price Index (PI).

In the work [14], the task of modeling the frequency of losses in insurance through actuarial calculations and the application of Bayesian "Classification and regression trees" (CART) models for accident frequency is addressed. Among many statistical methods, a result presented in [15] is highlighted, which, based on a publicly available database of workplace accidents and assuming that accident statistics information is definitive, ensures its statistical significance.

Scientific results from domestic scientists [16-17] can also be cited. The work [16] presents data on the dynamics of industrial injuries and occupational diseases in the mining industry of the Republic of Kazakhstan. The study [17] applies methods of economic and statistical analysis in the field of occupational safety and provides recommendations for optimizing the system of statistical monitoring of occupational safety. This includes the introduction of new methods for data collection and processing, updating forms of statistical reporting, and implementing proactive approaches to improving working conditions and preventing workplace accidents.

In this study, the object of scientific research is the occupational risk indicator, which is conducted based on the analysis of Kazakhstan statistics on injuries and mortality related to work activities and occupational diseases to assess occupational risk in detailed breakdown by TEA with a non-zero indicator of industrial injuries.

Various approaches are used for classification. Specifically, Hoffmann E, Chamie M. [18] distinguish between different types of classifications, i.e., "reference," "derived," or "related" classifications. Reference classifications result from international agreements approved by the United Nations Statistical Commission or another competent intergovernmental council. ISCO (International Standard Classification of Occupations, ILO) [19] and ISIC (International Standard Industrial Classification of All Economic Activities, UN) [7] are reference classifications and are recognized as such in the family of international economic classifications adopted at the second meeting of the Expert Group on International Classifications [20].

For a more detailed statistical analysis, all economic activities were considered at the five-digit code level (level) according to the OKED (Classification of Economic Activities of the Republic of Kazakhstan) NK RK-2019, the total number of which is 1018 types. Of these, economic activities with a non-zero list number of employees (hereinafter referred to as LNE), including small enterprises acting as a control group for calculating the indicators of industrial injuries on average for the analyzed period, amounted to 944 or 92.6%. Of these 944 types, in 504, there were registered incidents of accidents at work, which constitutes about 55% of the total number of economic activities with a non-zero list number of workers. As of the end of 2022, according to the BNS, the LNE amounted to 5,522.5 thousand people, with the share of workers employed in industry, construction, and the transport sector constituting 28.6%. In these sectors of the economy, the proportion of injured and deceased accounted for approximately 70% of the total number of workers who suffered injuries or died across the country (69.5% and 74.0%, respectively). As of the end of 2022, according to the BNS, the LNE amounted to 5,522.5 thousand people, with the share of workers employed in industry, construction, and the transport sector constituting 28.6%. In these sectors of the economy, the proportion of injured and deceased accounted for approximately 70% of the total number of workers who suffered injuries or died across the country (69.5% and 74.0%, respectively).

For the statistical analysis of industrial injuries, occupational diseases, and mortality due to workplace accidents, the following indicators for the years 2018-2022 were used:

- list number of workers, including small enterprises.
- number of individuals injured in work-related accidents with a loss of working capacity for 1 working day or more.
- number of fatalities in work-related accidents.
- number of workers employed in harmful and (or) hazardous working conditions (HHWC).

The aggregate values for the LNE and the number of workers who were injured or died in enterprises of the Republic of Kazakhstan for the period from 2018 to 2022, across 19 industries where incidents of workplace accidents are registered, are provided in Table 1.

Table 1 - Aggregate data by sectors of the Republic of Kazakhstan, averaged for 2018-2022.

Code and Name of the Sector	LNE (thous. people)	specific weight, %	number of injured, (people)	specific weight, %	including fatalities (people)	specific weight, %
A - Agriculture, forestry and fishing	144,3	2,6%	323	3,0%	59	6,0%
B - Mining and quarrying	223,5	4,0%	2 197	20,2%	135	13,7%
C - Manufacturing	471,2	8,5%	2 890	26,5%	167	16,9%
D - Electricity, gas, steam and air conditioning supply	114,8	2,1%	399	3,7%	67	6,8%
E - Water supply; sewerage, waste management and remediation activities	60,0	1,1%	272	2,5%	46	4,7%
F - Construction	398,9	7,2%	1 082	9,9%	233	23,6%
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	658,9	11,9%	448	4,1%	39	3,9%
H - Transportation and storage	311,3	5,6%	728	6,7%	84	8,5%
I - Accommodation and food service activities	94,6	1,7%	113	1,0%	5	0,5%
J - Information and communication	134,4	2,4%	74	0,7%	3	0,3%
K - Financial and insurance activities	114,1	2,1%	38	0,3%	1	0,1%
L - Real estate activities	106,6	1,9%	77	0,7%	8	0,8%
M - Professional, scientific and technical activities	209,9	3,8%	142	1,3%	20	2,0%
N - Administrative and support service activities	261,6	4,7%	475	4,4%	33	3,3%
O - Public administration and defence; compulsory social security	410,1	7,4%	255	2,3%	39	3,9%
P - Education	1089,2	19,7%	273	2,5%	12	1,2%
Q - Human health and social work activities	488,8	8,9%	977	9,0%	27	2,7%
R - Arts, entertainment and recreation	116,0	2,1%	75	0,7%	7	0,7%
S - Other service activities	114,3	2,1%	48	0,4%	4	0,4%
TOTAL for the Republic of Kazakhstan	5522,5	100,0	10 886	100	989	100

TEA are classified into occupational risk classes based on the following indicators:

1. Share of employment in harmful and (or) hazardous working conditions (K_1):

$$K_1 = \frac{N_{hhwc}}{LNE} \times 100\% \quad (1)$$

were,

N_{hhwc} - number of workers employed in harmful and (or) hazardous working conditions (HHWC);

LNE – list number of employees.

2. Total Incident Frequency Rate per 1 000 workers, TIFR (K_2);

$$K_2 = \frac{N_{acc}}{LNE} \times 1000 \quad (2)$$

were,

N_{acc} – number of accidents related to labor activity with loss of working capacity for 1 working day or more.

3. Fatality Incident Frequency Rate per 1 000 workers, FIFR (K_3);

$$K_3 = \frac{N_{fi}}{LNE} \times 1000 \quad (3)$$

were,

N_{fi} – number of deaths in labor-related accidents.

4. K_4 - ratio of K_3 and K_2 coefficients, FIFR /TIFR;

5. Distribution of the number of people affected and works in HHWC by type of economic activity (K_5):

$$K_5 = \frac{N_{acc\&hhwc(TEA)}}{N_t} \times 100\% \quad (4)$$

were,

$N_{acc\&hhwc(TEA)}$ - number of accidents related and works in HHWC by certain types of economic activity;

N_t – total number of accidents related and works in HHWC.

K_2 (frequency rate of all occupational accidents per 1,000 workers) was used as a key indicator. Based on these indicators, a new approach to classify TEA into 22 classes of occupational risk has been developed.

It should be noted here that to properly assess the injury rate, it is very important to analyse not only the frequency of fatal and non-fatal workplace injuries, but also to assess the ratio between them, which the ILO uses as an indicator of the level of non-fatal workplace injury underreporting. Fatal accidents are difficult to conceal, so data on such accidents are considered statistically reliable, unlike data on non-fatal accidents. Jukka Takala notes in his research paper that because there is no exaggeration of injury rates in reported data, "a low rate indicates underreporting of minor accidents, which are quite common"[21].

The results of the statistical analysis carried out are described below.

The figure below illustrates the values of indicator K_1 , i.e. the share of those employed in hazardous labor conditions by certain TEA with the industry code according to ISIC-Rev.4.

The industries with the largest number of workers employed in harmful and other unfavorable working conditions were manufacturing (27.8%), mining and quarrying (23.0%), transport and warehousing, and water supply (13.8% and 13.6% respectively).

Table 2 presents the estimated values of indicators regarding occupational injuries, namely TIFR, FIFR and the ratio of FIFR to TIFR.

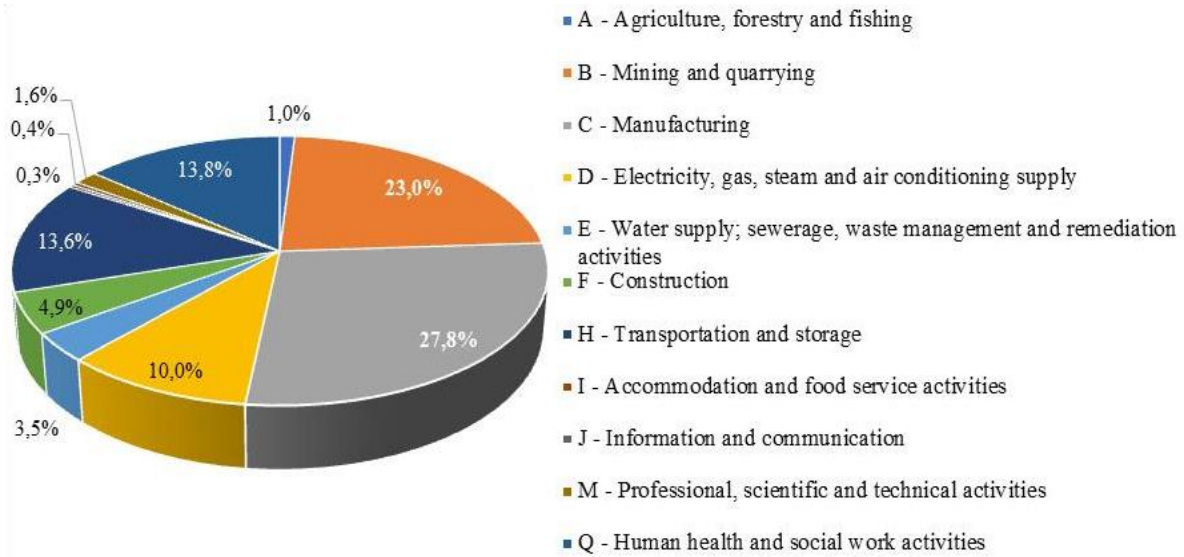


FIGURE 1 – Figure 1 - Specific weight of those employed in harmful labor conditions by certain types of economic activity (calculated based on official data of the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of RK, form 1-T (labor conditions) - Report on the number of workers employed in harmful and other unfavorable working conditions).

Table 2 - Indicators of occupational injuries by individual types of economic activity

Code and name of sectors	Indicators		
	K_2	K_3	K_4
Total for RK	0.40	0.04	0.10
A - Agriculture, forestry and fishing	0.47	0.08	0.18
B - Mining and quarrying	1.84	0.14	0.08
C - Manufacturing	1.19	0.09	0.07
D - Electricity, gas, steam and air conditioning supply	0.68	0.13	0.18
E - Water supply; sewerage, waste management and remediation activities	1.31	0.22	0.16
F - Construction	0.61	0.13	0.22
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	0.15	0.01	0.08
H - Transportation and storage	0.51	0.07	0.14
I - Accommodation and food service activities	0.21	0.01	0.05
J - Information and communication	0.19	0.01	0.04
K - Financial and insurance activities	0.08	0.00	0.02
L - Real estate activities	0.16	0.02	0.14
M - Professional. scientific and technical activities	0.18	0.03	0.15
N - Administrative and support service activities	0.22	0.02	0.10
O - Public administration and defence; compulsory social security	0.14	0.02	0.17
P - Education	0.05	0.00	0.06
Q - Human health and social work activities	0.39	0.01	0.04
R - Arts. entertainment and recreation	0.15	0.01	0.05
S - Other service activities	0.12	0.01	0.08

The level of occupational injuries is significantly differentiated by industries. The top industries in terms of these indicators are as follows:

1. Mining and quarrying ($K_2 - 1.84, K_3 - 0.14, K_4 - 0.08$).
2. Water supply; sewerage, waste management and remediation activities ($K_2 - 1.31, K_3 - 0.22, K_4 - 0.16$).
3. Manufacturing ($K_2 - 1.19, K_3 - 0.09, K_4 - 0.07$);
4. Electricity, gas, steam, and air conditioning supply ($K_2 - 0.68, K_3 - 0.13, K_4 - 0.18$).
5. Construction ($K_2 - 0.61, K_3 - 0.13, K_4 - 0.22$).

The assignment of TEA to occupational risk classes was made in accordance with the following six stages.

At the first step, four indicators were calculated for the five-digit level of classification of economic activities based on the arithmetic mean of statistical indicators for the previous five years.

Thus, 660 numerical values were found for 132 TEA $\{k_i^s\}$, (where $i = 1, \dots, 5; s = 1, \dots, 132$) which were then subjected to comparative procedures.

At the second step, the total set of economic activities subject to classification to occupational risk classes is determined on the basis of all sub-activities of economic activities according to the five-digit level of classification, dominating by the specific weight of victims in the related type of economic activity (Indicator K_2 values of which exceed the republican indicator -0.40), that is $k_2^s > k_2$, (where $i = 1, \dots, 5; s = 1, \dots, 132, k_2$ - republican indicator).

At the third step, the indicators for every TEA of the total population were assigned ratings in ascending order of values: the lowest value (zero) has a rate of 1, the highest value - a rate equal to the number of different values of the corresponding indicator for all types of economic activity (equal values were assigned the same rate); for K_1 - 1 to 89; K_2 - 1 to 77; K_3 - 1 to 87; K_4 - 1 to 94, $\max k_1^s = 89$, $\max k_2^s = 77$, $\max k_3^s = 87$, $\max k_4^s = 94$.

At the fourth step, weighted average values of ranks and their total value were calculated for every TEA of the total population.

At the fifth step the intervals were determined on the basis of quantitative (ordering the data from the minimum to the maximum, taking into account the average value for the entire population) and qualitative (equal to the total number of occupational risk classes) features of grouping (for indicators K_1, K_2, K_3, K_4 the minimum total value is 0.09 and the maximum is 3.14, respectively for occupational risk classes 1 and 22, for indicators K_1 and K_2 respectively 0.02 and 1.94).

At the sixth step, all TEA are classified by occupational risk classes according to the belonging of their total value of weighted average values of ranks to the corresponding interval value.

The application of this approach to the task of classifying TEA into classes of occupational risk based on a statistical study of indicators of occupational injuries and working conditions allowed 132 types (two-digit, including some five-digit) to be classified into 22 classes. In particular, the first class of occupational risk includes two-digit TEA of such industries as information and communication, financial and insurance activities, activities of households and extraterritorial organisations and bodies. At the same time, with the increase of the occupational risk class, the detailing by TEA increased, so the 22nd class (the highest) includes five-digit types of economic activity 07101 "Underground mining of iron ore" and 43991 "Mine construction", with the allocation of the two-digit code of mining and construction industry, respectively.

All the results obtained in the course of the calculation and analytical work on assigning TEA to classes have been introduced into the domestic labor legislation and are reflected in the Regulations for assigning TEA to classes of occupational risk, approved by Order No. 86 of the Minister of Labor and Social Protection of the Republic of Kazakhstan dated 5 March 2022.

CONCLUSIONS

Identification of risks of production factors for safety and health of workers and their further allocation to individual TEA requires priority attention. Classification of TEA into classes of occupational risk is based on statistical data on working conditions and occupational injuries using methods of mathematical analysis. Classification of occupational risks by individual activities should meet the criteria of similarity of risk characteristics in one group of industry affiliation, the coverage of statistical data should be large enough for data reliability, and the definitions of each class should be sufficiently accurate to exclude the possibility that two identical risks will be classified differently. Therefore, the reliability and transparency of such information will subsequently have a meaningful impact on the outcome of the work undertaken. A scientifically substantiated methodology for classifying TEA into occupational risk classes will serve as a reference point for applying differentiated insurance rates for industrial accident insurance.

ACKNOWLEDGMENTS

The statistical database provided by the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan was used in the course of the research study. In this regard, the authors express their gratitude for assistance in collecting the necessary information.

Scientific results were obtained within the framework of scientific and technical program №BR11965728 "Economic problems of safe labor and institutional transformation of the insurance mechanism in the Republic of Kazakhstan" under the program-target financing of the Ministry of Labor and Social Protection of Population of the Republic of Kazakhstan.

References

- 1 Трудовой кодекс Республики Казахстан от 23 ноября 2015 года № 414-V ЗПК. [Электронный ресурс]. -URL: <https://adilet.zan.kz/rus/docs/K1500000414>.
- 2 Приказ Министра труда и социальной защиты населения Республики Казахстан от 5 марта 2022 года № 86 "Об утверждении Правил отнесения видов экономической деятельности к классам профессионального риска". [Электронный ресурс]. -URL: <https://adilet.zan.kz/rus/docs/V2200027085>.
- 3 Приказ Комитета технического регулирования и метрологии Министерства индустрии и новых технологий РК № 683-од от 14.12.2007 г. "Общий классификатор видов экономической деятельности//ОКЭД НК РК 03-2019 пятизначный". [Электронный ресурс]. -URL: https://online.zakon.kz/Document/?doc_id=33987425.
- 4 Закон Республики Казахстан от 7 февраля 2005 года № 30 "Об обязательном страховании работника от несчастных случаев при исполнении им трудовых (служебных) обязанностей". [Электронный ресурс]. -URL: <https://adilet.zan.kz/rus/docs/Z050000030>.
- 5 Официальные данные Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан. [Электронный ресурс]. -URL: <https://www.stat.gov.kz/official/industry/25/statistic/6>.
- 6 Официальные данные Национального банка Республики Казахстан. Сводный отчет о страховых премиях по страховым (перестраховочным) организациям. Сводный отчет о страховых выплатах по страховым (перестраховочным) организациям. [Электронный ресурс]. -URL: https://www.nationalbank.kz/ru/news/svodnyye_otchet_ye_otchet_ye_str_sektor.
- 7 Международная стандартная отраслевая классификация всех видов экономической деятельности. [Электронный ресурс]. -URL: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_russian.pdf.
- 8 Павленко Д. Введение в машинное обучение и искусственные нейронные сети [Электронное издание]. https://foobar167.github.io/page/vvedeniye_v_mashinnoye_obucheniye_i_iskusstvennyye_neyronnyye_seti.html.
- 9 Воронцов К.В. Математические методы обучения по прецедентам (теория обучения машин). [Электронный ресурс]. -URL: http://www.machinelearning.ru/wiki/images/6/6d/voron_ml-1.pdf.
- 10 Об утверждении Правил управления профессиональными рисками. Приказ Министра труда и социальной защиты населения Республики Казахстан от 11 сентября 2020 года № 363 [Электронный ресурс]. -URL: <https://adilet.zan.kz/rus/docs/V2200027085>.
- 11 Jinxian Weng, Xiafan Gan, Zheyu Zhang, A quantitative risk assessment model for evaluating hazard transportation accident risk // Safety Science. -2021. -Vol. 137. P. 105198. -URL: <https://doi.org/10.1016/j.ssci.2021.105198>.
- 12 Maryam Sadat Hajakbari, Behrouz Minaei-Bidgoli, A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data // Journal of Loss Prevention in the Process Industries. -2014. -Vol. 32. -P. 443-453. -URL: <https://doi.org/10.1016/j.jlp.2014.10.013/>.
- 13 Kamran Gholamizadeh, Esmail Zarei, Mohammad Yazdi, Matilde A. Rodrigues, Nasrin shirmohammadi-Khorram, Iraj Mohammadfam. An integration of intelligent approaches and economic criteria for predictive analytics of occupational accidents // Decision Analytics Journal. -2023. -Vol. 9. P. 100357. -URL: <https://doi.org/10.1016/j.dajour.2023.100357>.
- 14 Yifan Huang, Shengwang Meng. A Bayesian nonparametric model and its application in insurance loss prediction // Insurance: Mathematics and Economics. 2020. -Vol. 93. -P. 84-94. DOI: <https://doi.org/10.1016/j.insmatheco.2020.04.010>.
- 15 Tuula Rasanen, Arto Reiman, Kai Puolamaki, Rafael Savvides, Emilia Oikarinen, Eero Lantto, Finding statistically significant high accident counts in exploration of occupational accident data // Journal of Safety Research. -2022. -Vol. 82. -P. 28-37. -URL: <https://doi.org/10.1016/j.jsr.2022.04.003>.
- 16 Yerdessov N, Izdenov A, Beisenov T, Suleimenova R, Serik B, Sraubaev E. Industrial traumatism and occupational morbidity in mining industry of Kazakhstan // Journal Public Health Res. -2021. 24;11(1):2169.

- 17 Abikenova S., Aitimova S., Daumova G., Koval A., Sarybayeva I. Statistical Monitoring of OSH: Analysis of Deviations and Recommendations for Optimization // International Journal of Safety and Security Engineering. -2023. -Vol. 13(6). -P. 1049-1059.
- 18 Hoffmann E, Chamie M. Standard Statistical Classifications: Basic Principles. -URL: <http://www.un.org/depts/unsd/statcom/principl.pdf>, 1999.
- 19 Международная стандартная классификация занятий 2008 г. [Электронный ресурс]. -URL: <https://www.ilo.org/wcmsp5/groups/public/europe/rogeneva/sromoscow/documents/publication/wcms306603.pdf>.
- 20 Krieger N, Williams DR, Moss NE. Measuring social class in US public health research// Annu Rev Public Health 1997; 18:341-78.
- 21 Takala J. Global Estimates of Fatal Occupational Accidents// Epidemiology. -1999; Sep; 10(5):640-6.

Ш.К. Әбікенова¹, Р. Марчеллони², Ш.Т.Әйтiмова³

¹ "Рагуз" жұмыс берушілер ұлттық конфедерациясы", әл-Фараби даңғ, 77/8, Алматы, Қазақстан

² Өндірістегі жазатайым оқиғалардан сақтандыру ұлттық институты (INAIL), le Pastore ave., 6, Рим, Италия

³ Еңбек және халықты әлеуметтік қорғау министрлігінің Республикалық еңбекті қорғау бойынша ғылыми-зерттеу институты, Кравцова, 18, Астана, Қазақстан

Кәсіби тәуекелдің статистикалық көрсеткіштер негізінде жіктелуінің математикалық аспектілері

Аннотация: Кәсіптік тәуекел Қазақстан Республикасындағы әлеуметтік-еңбек қатынастарын реттеудегі салыстырмалы түрде жаңа ұғым болып табылады. Еңбек қызметіне байланысты жазатайым оқиға нәтижесінде еңбек (қызметтік) міндеттерін атқару кезінде еңбекке қабілеттілігінен айырылу не қызметкердің қайтыс болу тәуекелінің көрсеткіші ретінде қолданылады. Экономикалық қызметтің 132 түрінде (екі таңбалы, соның ішінде кейбір бес таңбалы) үлкен деректер бар Машиналық оқыту теориясын қолдану кәсіби тәуекелдің 22 санатына бөлінген жіктеу процедурасын жүргізуге мүмкіндік берді. Кәсіби тәуекелді бағалаудың әдістемелік негізі сенімді статистикалық деректерді пайдалану арқылы салыстырудың, топтастырудың және жүйелеудің жалпы қабылданған статистикалық әдістері болып табылады.

Зерттеудің мақсаты - қызметкерді еңбек міндеттерін орындау кезінде жазатайым оқиғалардан міндетті сақтандыру жүйесінде кәсіптік тәуекел санаттары бойынша экономикалық қызмет түрлерін жіктеудің математикалық тәсілдерінің бірін қолдану. Бұл тәсілді пайдалану экономикалық қызметтің 132 түрін (екі таңбалы, оның ішінде кейбір бес таңбалы) кәсіптік тәуекелдің 22 санатына жатқызуға мүмкіндік берді. Кәсіби санатты арттыру кезінде экономикалық қызмет түрлерін бес таңбалы жіктеуді және егжей-тегжейлі көрсетуді енгізу қажеттілігі анықталды. Сондықтан 22 (жоғары) сыныпқа тау-кен өнеркәсібі мен құрылыс саласының екі таңбалы кодының 07101 "Темір кендерін жерасты өндіру" және 43991 "Шахта құрылысы" бес таңбалы экономикалық қызмет түрлері кірді. Ғылыми нәтижелер Қазақстан Республикасы Еңбек және халықты әлеуметтік қорғау министрлігінің бағдарламалық-нысаналы қаржыландыру шеңберінде алынды (№ BR11965728 "Қазақстан Республикасында қауіпсіз еңбектің экономикалық проблемалары және сақтандыру механизмінің институционалдық қайта құрулары").

Түйін сөздер: кәсіптік тәуекел, еңбек статистикасы, жазатайым оқиға жиілігі, өндірістік жарақат деңгейі, тәуекелділік (рангтар) бойынша сұрыптау, ауыспалы диапазон.

Ш.К. Абикенова¹, Р.Марчеллони², Ш.Т. Айтiмова³

¹ Национальная конфедерация работодателей "Рагуз", пр. аль-Фараби, 77/8, Алматы, Казахстан

² Национальный институт страхования от несчастных случаев на производстве (INAIL), le Pastore, 6, Рим, Италия

³ Республиканский научно-исследовательский институт охраны труда Министерства труда и социальной защиты населения Республики Казахстан, Кравцова, 18, Астана, Казахстан

Математические аспекты профессионального риска и его классификации с учетом статистических показателей

Аннотация: Профессиональный риск является относительно новым понятием в регулировании социально-трудовых отношений в Республике Казахстан и применяется в качестве индикатора риска утраты трудоспособности либо смерти работника при выполнении трудовых (служебных) обязанностей в результате несчастного случая, связанного с трудовой деятельностью. В статье раскрываются статистические аспекты содержания профессионального риска в подтверждении математической вездесущности и обоснованности научного аппарата. Методической основой оценки профессионального риска являются общепринятые статистические методы сравнения, группировки и систематизации посредством сопоставления достоверных статистических данных.

Цель исследования состоит в применении одного из математических подходов к распределению видов экономической деятельности по классам профессионального риска в системе обязательного страхования работника от несчастных случаев при выполнении им трудовых обязанностей. Применение теории Машинного обучения с большими данными в 132 видах экономической деятельности (двухзначные, включая некоторые пятизначные) позволило произвести процедуру классификации с разбивкой на 22 класса профессионального риска. Была выявлена необходимость при повышении класса профессионального введения пятизначной классификации и детализации видов экономической деятельности, так в 22 класс (самый высокий) вошли пятизначные ВЭД 07101 "Добыча железных руд подземным способом" и 43991 "Строительство шахт", с выделением из двухзначного кода горнодобывающей промышленности и строительной отрасли соответственно. Научные результаты получены в рамках программно-целевого финансирования Министерством труда и социальной защиты населения

Л.Н. Гумилев атындағы ЕҰУ Хабаршысы. Математика. Компьютерлік ғылымдар. Механика, 2023, Том 145, №4

Вестник ЕНУ им. Л.Н. Гумилева. Математика. Компьютерные науки. Механика, 2023, Том 145, №4

Республики Казахстан (научно-техническая программа №BR11965728 "Экономические проблемы безопасного труда и институциональные преобразования механизма страхования в Республике Казахстан").

Ключевые слова: профессиональный риск, статистика труда, частота несчастных случаев, уровень производственного травматизма, сортировка по релевантности риска (ранги), вариативный диапазон.

References

- 1 Trudovoj kodeks Respubliki Kazahstan ot 23 noyabrya 2015 goda № 414-V ZRK. Labor code of the Republic of Kazakhstan [Code of the Republic of Kazakhstan dated 23 November, 2015 no. 414- V ZRK]. [Electronic resource]. Available at: <https://adilet.zan.kz/rus/docs/K1500000414>.
- 2 Prikaz Ministra truda i social'noj zashchity naseleniya Respubliki Kazahstan ot 5 marta 2022 goda № 86 "Ob utverzhdenii Pravil otneseniya vidov ekonomicheskoy deyatel'nosti k klassam professional'nogo riska" [Order of the Minister of Labor and Social Protection of the Population of the Republic of Kazakhstan dated 5 March 2022 No. 86 "On Approval of the Regulations for Attributing Types of Economic Activities to Classes of Occupational Risk"]- [Electronic resource]. Available at: <https://adilet.zan.kz/rus/docs/V2200027085>.
- 3 Prikaz Komiteta tekhnicheskogo regulirovaniya i metrologii Ministerstva industrii i novyh tekhnologij RK № 683-od ot 14.12.2007 g. "Obschij klassifikator vidov ekonomicheskoy deyatel'nosti//OKED NK RK 03-2019 pyatiznachnyj"[Order of the Committee for Technical Regulation and Metrology of the Ministry of Industry and New Technologies of the Republic of Kazakhstan No. 683-od dated 14.12.2007. "General Classifier of Economic Activities. OKED NK RK 03-2019 five-digit"]. Available at: https://online.zakon.kz/Document/?doc_id=33987425.
- 4 Zakon Respubliki Kazahstan ot 7 fevralya 2005 goda № 30 "Ob obyazatel'nom strahovanii rabotnika ot neschastnyh sluchaev pri ispolnenii im trudovyh (sluzhebnyh) obyazannostej" [Act No. 30 of the Republic of Kazakhstan of 7 February 2005 "On compulsory insurance of workers against accidents in the performance of their employment (official) duties"]. [Electronic resource]. Available at: <https://adilet.zan.kz/rus/docs/Z050000030>.
- 5 Oficial'nye dannye Byuro nacional'noj statistiki Agentstva po strategicheskomu planirovaniyu i reformam Respubliki Kazahstan. Official data of the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan (eng). [Electronic resource]. Available at: <https://www.stat.gov.kz/oficial/industry/25/statistic/6>.
- 6 Oficial'nye dannye Nacional'nogo banka Respubliki Kazahstan. Svodnyj otchet o strahovyh premiyah po strahovym (perestrahovochnym) organizacijam. Svodnyj otchet o strahovyh vyplatah po strahovym (perestrahovochnym) organizacijam. [Official data of the National Bank of the Republic of Kazakhstan. Summary report on insurance premiums for insurance (reinsurance) organizations. Summary report on insurance payments for insurance (reinsurance) organizations][Electronic resource]. Available at: https://www.nationalbank.kz/ru/news/svodnye_otchety_str_sektor.
- 7 Mezhdunarodnaya standartnaya otraslevaya klassifikaciya vsekh vidov ekonomicheskoy deyatel'nosti. [International Standard Industrial Classification] - [Electronic resource]. Available at: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_Russian.pdf.
- 8 Paulenka D. Vvedenie v mashinnoe obuchenie i iskusstvennye neyronnye seti [Introduction to Machine Learning and Artificial Neural Networks] [Electronic resource]. <https://foobar167.github.io/page/vvedeniye-v-mashinnoye-obuchenie-i-iskusstvennyye-neyronnyye-seti.html>.
- 9 Voroncov K. V. Matematicheskie metody obucheniya po precedentam (teoriya obucheniya mashin) [Mathematical methods of learning from precedents (machine learning theory)]. [Electronic resource]. Available at: http://www.machinelearning.ru/wiki/images/6/6d/voron_ml-1.pdf.
- 10 Ob utverzhdenii Pravil upravleniya professional'nymi riskami. Prikaz Ministra truda i social'noj zashchity naseleniya Respubliki Kazahstan [On Approval of the Regulations of Occupational Risk Management. Order of the Minister of Labor and Social Protection of the Population of the Republic of Kazakhstan] dated 11 September 2020 No. 363. [Electronic resource]. Available at: <https://adilet.zan.kz/rus/docs/V2200027085>.
- 11 Jinxian Weng, Xiafan Gan, Zheyu Zhang, A quantitative risk assessment model for evaluating hazmat transportation accident risk, Safety Science. 2020. Vol. 137. P. 105198. DOI: <https://doi.org/10.1016/j.ssci.2021.105198>.
- 12 Maryam Sadat Hajakbari, Behrouz Minaei-Bidgoli, A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data, Journal of Loss Prevention in the Process Industries. 2014. Vol. 32. P. 443-453. Available at: <https://doi.org/10.1016/j.jlp.2014.10.013/>.
- 13 Kamran Gholamzadeh, Esmail Zarei, Mohammad Yazdi, Matilde A. Rodrigues, Nasrin shirmohammadi-Khorram, Iraj Mohammadfam. An integration of intelligent approaches and economic criteria for predictive analytics of occupational accidents, Decision Analytics Journal. 2023. Vol. P. 100357. Available at: <https://doi.org/10.1016/j.dajour.2023.100357>.
- 14 Yifan Huang, Shengwang Meng. A Bayesian nonparametric model and its application in insurance loss prediction, Insurance: Mathematics and Economics. 2020. -Vol. 93. -P. 84-94. DOI: <https://doi.org/10.1016/j.insmatheco.2020.04.010>.

- 15 Tuula Rasanen, Arto Reiman, Kai Puolamaki, Rafael Savvides, Emilia Oikarinen, Eero Lantto, Finding statistically significant high accident counts in exploration of occupational accident data, Journal of Safety Research. 2022. Vol.82. P. 28-37. Available at: <https://doi.org/10.1016/j.jsr.2022.04.003>.
- 16 Yerdessov N, Izdenov A, Beisenov T, Suleimenova R, Serik B, Sraubaev E. Industrial traumatism and occupational morbidity in mining industry of Kazakhstan, J Public Health Res. 2021 Sep 24;11(1):2169. DOI: 10.4081/jphr.2021.2169. PMID: 34558882; PMCID: PMC8874858.
- 17 Abikenova, S., Aitimova, S., Daumova, G., Koval, A., Sarybayeva, I. Statistical Monitoring of OSH: Analysis of Deviations and Recommendations for Optimization, International Journal of Safety and Security Engineering. 2023. 13(6). P. 1049-1059.
- 18 Hoffmann E, Chamie M. Standard Statistical Classifications: Basic Principles. 1999. Available at: <http://www.un.org/depts/unsd/statcom/principi.pdf>.
- 19 International Standard Classification of Occupations 2008. [Electronic resource]. Available at: <https://www.ilo.org/wcmsp5/groups/public/europe/ro-geneva/sromoscow/documents/publication/wcms306603.pdf/>.
- 20 Krieger N, Williams DR, Moss NE. Measuring social class in US public health research, Annu Rev Public Health. 1997. Vol. 18. P. 341-78.
- 21 Takala J. Global Estimates of Fatal Occupational Accidents, Epidemiology. 1999. Sep. 10(5). P. 640-6.

Information about authors:

Әбікенова Шолпан Кәкімжанқызы - байланыс үшін автор, ф.-м.ғ.к., доцент, "PARYZ" жұмыс берушілер ұлттық конфедерациясы" ММ бас директорының орынбасары, әл-Фараби даңғ., 77/8, Алматы, Қазақстан.

Раффаэлло Марчеллони - статистика, демографиялық және актуарлық ғылымдар магистрі, Статистика және актуарлық есептеулер департаментінің білікті актуарийі, Өндірістегі жазатайым оқиғалардан сақтандырудың ұлттық институты (INAIL), le Pastore ave., 6, Рим, Италия.

Әйтимова Шынар Тұрсынханқызы - экономика ғылымдарының магистрі, Еуразия ұлттық университетінің докторанты, Еңбек және халықты әлеуметтік қорғау министрлігінің Республикалық еңбекті қорғау ғылыми-зерттеу институты экономикалық өлшемдер және статистика бөлімінің статистика секторының меңгерушісі. Қазақстан Республикасы, Кравцова, 18, Астана, Қазақстан.

Sholpan Abikenova - corresponding author, cand. of phys.-math. sci., Associate Professor, Deputy General Director, ALE "National Confederation of Employers "PARYZ", al-Farabi ave., 77/8, Almaty, Kazakhstan.

Raffaello Marcelloni - Master's Degree in Statistical, Demographic and Actuarial Sciences, qualified actuary in General Statistical and Actuarial department, National Institute for Insurance against Accidents at Work (INAIL), le Pastore, 6, Roma, Italy.

Shynar Aitimova - Master's Degree of Economics, doctoral student of the Eurasian National University, head of the statistics sector of the department of economic measurements and statistics, Republican Research Institute of Labor Protection of the Ministry of Labor and Social Protection of the Population of the Republic of Kazakhstan, Kravtsova, 18, Astana, Kazakhstan.

Received 07.12.2023

Бас редактор:

Н. Темірғалиев

Жауапты редактор:

А.Ж. Жұбанышева

Л.Н. Гумилев атындағы Еуразия ұлттық университетінің
хабаршысы. Математика. Компьютерлік ғылымдар. Механика сериясы.
- 2023. 4(145)- Астана: ЕҰУ. 35-б. Басуға қол қойылды: 29.12.2023.
Ашық қоланыстағы электронды нұсқа: <http://bulmathmc.enu.kz/>

Авторларға арналған нұсқаулықтар, публикациялық этика журнал сайтында берілген: <http://bulmathmc.enu.kz/>

Мазмұнына типография жауап бермейді

Редакция мекен-жайы: 010008, Қазақстан Республикасы, Астана қ.,
Сәтпаев көшесі, 2.

Л.Н. Гумилев атындағы Еуразия ұлттық университеті
Тел.: +7(7172) 70-95-00 (ішкі 31-410)

Л.Н. Гумилев атындағы Еуразия ұлттық университетінің баспасында басылды