ISSN 3007-0155
eISSN 3007-0163

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**

№1(150)/2025

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, 2025
Astana, 2025

БАС РЕДАКТОРЫ

Теміргалиев 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*
Алимхан Килан	*PhD, проф., Л.Н. Гумилев ат. ЕҰУ, Астана, Қазақстан*
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*
Нуртазина К.Б.	*ф.-м.ғ.к., доцент, Л.Н. Гумилев атындағы ЕҰУ, Астана, Қазақстан*
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*
Тауғынбаева Г.Е.	*PhD, Л.Н. Гумилев ат. ЕҰУ, Астана, Қазақстан*
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).

© Л.Н. Гумилев атындағы Еуразия ұлттық университеті

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*
Alexander Iosevich	*PhD, Prof., University of Rochester, New York, USA*
Alimhan Keylan	*PhD, Prof., L.N. Gumilyov ENU, Astana, Kazakhstan*
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 Communications and Informatics, Moscow, Russia*
Karlygash Nurtazina	*Cand. of Phys.-Math. Sci., Assoc.Prof., L.N. Gumilyov ENU, Astana, Kazakhstan*
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*
Galiya Taugynbayeva	*PhD, L.N. Gumilyov ENU, Astana, Kazakhstan*
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*
Алимхан Килан	*PhD, проф., ЕНУ имени Л.Н.Гумилева, Астана, Казахстан*
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*
Нуртазина К.Б.	*к.ф.-м.н., доцент, ЕНУ имени Л.Н.Гумилева, Астана, Казахстан*
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*
Таугынбаева Г.Е.	*PhD, ЕНУ имени Л.Н.Гумилева, Астана, Казахстан*
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).

© Евразийский национальный университет имени Л.Н. Гумилева

Л.Н. Гумилев атындағы Еуразия ұлттық университетінің хабаршысы. Математика, компьютерлік ғылымдар, механика сериясы, №1(150)/2025

Bulletin of L.N. Gumilyov Eurasian National University. Mathematics, computer science, mechanics series, №1(150)/2025

Вестник Евразийского национального университета имени Л.Н.Гумилева. Серия Математика, компьютерные науки, механика, №1(150)/2025

МАЗМҰНЫ
CONTENTS
СОДЕРЖАНИЕ

Умбет С., Кожирбаев Ж.М. Қазақ тіліне арналған Wav2Vec2 моделін пайдалана отырып, ауызша сөйлеу тілін сәйкестендіру
Umbet S., Kozhirbayev Zh.M. Identification of the spoken language using the Wav2Vec2 model for the Kazakh language
Умбет С., Кожирбаев Ж.М. Идентификация языка устной речи с использованием модели Wav2Vec2 для казахского языка 6

Наурызбаев Н., Шоманова А., Темірғалиев Н. Функцияларды тригонометриялық Фурье қатарларының мүшелерінің ақырлы қосындылары арқылы жуықтаудың Банах өлшемі бойынша орташа квадраттық қателіктері
Nauryzbaev N., Shomanova A., Temirgaliyev N. Average Square Errors by Banach Measure of Recovery of Functions by Finite Sums of Terms of Their Trigonometric Fourier Series
Наурызбаев Н., Шоманова А., Темиргалиев Н. Средние квадратические погрешности по мере Банаха восстановления функций конечными суммами членов их тригонометрических рядов Фурье 17

Әбікенова Ш., Айтимова Ш., Кочербаева А. Еңбек қорғау шығындарын есепке алу және талдау бойынша ақпараттық-талдамалық деректер базасын қалыптастырудың негізі мен ғылыми дәйегі ретіндегі математикалық әдістер
Abikenova Sh., Aitimova Sh., Kocherbaeva A. Mathematical methods as a basis and scientific justification for the development of an information-analytical database for the accounting and analysis of OSH expenditures
Абикенова Ш., Айтимова Ш., Кочербаева А. Математические методы как основа и научное обоснование к формированию информационно-аналитического банка данных по учету и анализу затрат на охрану труда 25

Bulletin of L.N. Gumilyov Eurasian National University. Mathematics, computer science, mechanics series, 2025, Vol. 150, №1, P. 6-16.
http://bulmathmc.enu.kz, E-mail: vest_math@enu.kz

Статья
МРНТИ: 28.23.17

ИДЕНТИФИКАЦИЯ ЯЗЫКА УСТНОЙ РЕЧИ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ WAV2VEC2 ДЛЯ КАЗАХСКОГО ЯЗЫКА [1]

Умбет С. [1], Кожирбаев Ж.М. [2]
[1] *Трирский университет, ул. Университетсринг 15, 54296, Трир, Германия*
[2] *National Laboratory Astana, пр. Кабанбай батыра, 53, Астана, Казахстан*
E-mail:zhanibek.kozhirbayev@nu.edu.kz

Аннотация. В данном исследовании представлена разработка и тонкая настройка модели идентификации языка устной речи с использованием варианта XLSR (Cross-Lingual Speech Recognition) Wav2Vec2. Обученная на богатом и разнообразном наборе данных, охватывающем шесть языков, с особым акцентом на языках с низкими ресурсами, таких, как казахский, модель демонстрирует замечательные возможности в многоязычном распознавании речи. Благодаря обширной оценке тонко настроенная модель не только превосходит существующие эталонные показатели, но и превосходит другие современные модели, включая варианты Whisper. Достигнув впечатляющего результата F1 92,9% и точности 93%, модель демонстрирует свою производительность в реальных многоязычных и малоресурсных сценариях. Эта работа вносит значительный вклад в развитие технологий распознавания речи, предоставляя надежное решение для идентификации языка в различных языковых средах, особенно в недостаточно представленных языковых условиях. Его успех подчеркивает потенциал моделей на основе Wav2Vec2 в улучшении систем обработки речи в многоязычных контекстах с низкими ресурсами. Итоги данного анализа могут способствовать разработке надежных и эффективных систем автоматического распознавания речи, оптимизированных для казахского языка. Такие технологии найдут применение в различных областях, включая преобразование речи в текст, работу голосовых ассистентов и инструменты голосовой коммуникации.

Ключевые слова: идентификации языка, идентификации языка устной речи, казахский язык, Wav2Vec2, XLSR.

DOI: https://doi.org/10.32523/bulmathenu.2025/1.1

2000 Mathematics Subject Classification: 68T10

1. ВВЕДЕНИЕ

Идентификация языка (LID) является важнейшей задачей в области обработки речи, выступая в качестве основы для многочисленных передовых приложений, таких, как автоматическое распознавание речи (ASR), перевод в реальном времени и виртуальные

[1]Работа выполнена при поддержке грантового финансирования проектов Комитета науки Министерства науки и высшего образования Республики Казахстан (грант No. AP23489529)

помощники, которые могут обслуживать многоязычные среды [1]. Способность систем обнаруживать и адаптироваться к разным языкам на лету является не только вопросом удобства, но и необходимостью в сегодняшнем все более глобализированном мире. Однако создание эффективных моделей LID далеко не просто, особенно при работе с языками, которые имеют ограниченные обучающие данные или ресурсы. Недавние достижения в области обработки речи сместились в сторону более сложных, эффективных с точки зрения данных подходов, которые используют кросс-языковое трансферное обучение. Одним из выдающихся методов в этом отношении стала разработка Wav2Vec2, которая произвела революцию в том, как обрабатываются речевые данные, путем обучения представлений непосредственно из необработанных данных [2]. Основываясь на этом, вариант XLSR Wav2Vec2 стал важным решением для задач многоязыковой идентификации языка. Изучая общие речевые представления на разных языках, XLSR оказался особенно эффективным для языков с низкими ресурсами — языков, где доступность обучающих данных ограничена, что затрудняет эффективное обучение моделей с использованием традиционных методов. Основное внимание в этой работе уделяется изучению того, как XLSR можно настроить для решения задач идентификации речи на разных языках. В частности, эта работа нацелена на шесть языков: казахский, русский, английский, итальянский, испанский и корейский. Стоит отметить, что казахский, один из основных языков в этом исследовании, часто недостаточно представлен в наборах данных идентификации языка, что делает его отличным кандидатом для проверки адаптивности модели к средам с низкими ресурсами. Включая сочетание языков с высоким ресурсом (например, английский, русский) и низким ресурсом (например, казахский), это исследование направлено на демонстрацию универсальности и надежности модели в реальной многоязычной среде. В этой работе не только описывается процесс обучения и метрики оценки, но и проводятся сравнения с возможностями LID существующей модели ASR, включая несколько вариантов модели Whisper. Эти сравнения необходимы для объективной оценки сильных сторон и ограничений тонко настроенной модели XLSR. Результаты этого исследования показали многообещающие результаты, особенно с точки зрения точности и оценок F1, что указывает на то, что тонко настроенная модель способна эффективно справляться с задачей идентификации многоязычного языка, особенно для языков с низким ресурсом, таких, как казахский. Основные вклады этого исследования включают:

1. Разработка и настройка модели на основе XLSR: Создание тонко настроенной модели, специально оптимизированной для задачи идентификации языков, включая казахский.
2. Сравнительный анализ с современными моделями: Проведение детального сравнения с существующими ASR-системами, включая несколько вариантов модели Whisper, для объективной оценки производительности.
3. Обоснование возможностей Wav2Vec2 для LID: Подтверждение эффективности использования модели Wav2Vec2 и ее вариантов, таких, как XLSR, для задач многоязычной идентификации речи.

Данная статья организована следующим образом: в разделе 2 представлено подробное обсуждение технических аспектов моделей Wav2Vec2.0 и XLSR, а также обзор литературы, связанной с идентификацией устной речи. Раздел 3 содержит детальное описание использованного набора данных, а также подходов, основанных на архитектуре XLSR, примененных в наших экспериментах. Результаты проведенных экспериментов изложены в разделе 4. Наконец, в разделе 5 подведены итоги исследования, представлены основные выводы, сделанные на основе экспериментов, и выделены перспективные направления для дальнейших исследований.

2. ОСНОВНАЯ ЧАСТЬ

В этом разделе представлен краткий обзор соответствующей литературы, связанной с этой статьей, разделенный на два подраздела: Wav2Vec2.0 и XLSR, а также идентификация языка устной речи.

Wav2Vec 2.0 представляет собой тщательно разработанную модель для расшифровки речи, встроенной в аудиосигналы, с применением самоконтролируемой методологии предварительного обучения. Эта методология позволяет извлекать полезные представления из больших объемов немаркированных аудиоданных. Модель объединяет ключевые принципы нескольких предшествующих подходов, включая Contrastive Predictive Coding (CPC) [3], Model Predictive Control (MPC) [4], wav2vec [5] и vq-wav2vec [6]. Архитектура Wav2Vec 2.0 интегрирует сверточные нейронные сети (CNN) и трансформеры, что позволяет эффективно обрабатывать как локальные особенности, так и глобальные шаблоны в аудиоданных. В основе модели лежит многоуровневый сверточный кодировщик признаков $f : X \to Z$, который преобразует необработанные аудиосигналы X в скрытые представления речи $z_1, \ldots, z_T$. Эти представления затем обрабатываются трансформерной сетью с использованием маскирования $g : Z \to C$, которая преобразует скрытые представления в дискретные выходные данные $q_1, \ldots, q_T$, используемые в качестве целевых значений в задаче самоконтролируемого обучения [7, 8]. Модуль трансформера контекстуализирует квантованные представления с помощью блоков внимания, создавая дискретные контекстуальные представления $c_1, \ldots, c_T$. Кодировщик признаков состоит из семи сверточных блоков с 512 каналами, шириной ядер {10, 3, 3, 3, 3, 2, 2} и шагами {5, 2, 2, 2, 2, 2, 2}. Трансформерная сеть включает 24 блока с размерностью 1024 и внутренней размерностью 4096, а также 16 головок внимания, обеспечивающих высокую степень контекстуализации.

XLSR [9] — многоязычная модель, созданная на основе межъязыковой модели XLM-R, предназначенная для решения задач многоязычной и межъязыковой обработки естественного языка (NLP). Она базируется на архитектуре Wav2Vec 2.0 и обладает уникальной способностью извлекать скрытые квантованные представления, охватывающие множество языков. Это достигается использованием метода квантования произведения, который выбирает квантованные представления из кодовых книг, а процесс выбора реализуется через дифференцируемую технику Gumbel-Softmax. Архитектура XLSR напоминает модель двунаправленного кодировщика представлений трансформера (BERT) [8], но имеет важное отличие: она включает 53 языковых вложения, что позволяет модели эффективно работать с каждым поддерживаемым языком. Этот подход даёт возможность модели улавливать тонкости языков, даже если они имеют схожее написание или произношение. Модель включает в себя 500 миллионов параметров, что делает её одной из крупнейших многоязычных моделей в мире. XLSR-53 была обучена на огромном корпусе данных, включающем тексты и речь более чем на 53 языках. Её способность к межъязыковому пониманию особенно ценна для трансферного обучения. Это позволяет использовать модель, обученную на одном языке, для достижения высоких результатов на другом языке с минимальными затратами на дополнительное обучение.

Идентификация языка стало важной областью исследований в области обработки речи [10, 11, 12, 13], с различными моделями, введенными для повышения точности, особенно в многоязычных средах. В одном из таких исследований [14] были применены модели глубокого обучения для распознавания устной речи с использованием сверточных нейронных сетей (CNN) на спектрограммах необработанных аудиосигналов. Этот подход показал многообещающие результаты с точностью 98% на тестовом наборе данных, что демонстрирует эффективность CNN в извлечении специфичных для языка признаков из речевых данных. В этой работе [15] представлен многомасштабный подход к извлечению признаков для повышения точности идентификации языка (LID) в сложных акустических средах. Заменив базовую сеть извлечения признаков архитектурой SE-Res2Net-CBAM-BILSTM, исследование достигло значительного улучшения производительности. Эксперименты, проведенные на многоязычном наборе данных коктейльной вечеринки, продемонстрировали надежность модели, при этом точность достигла 97,6% для набора данных восточного языка и 75% для имитированного набора данных коктейльной вечеринки. Кроме того, исследование подчеркивает эффективность фокальной потери в улучшении производительности модели, особенно при обработке сценариев несбалансированных данных. Недавние достижения в

области обнаружения языка для казахского и русского языков также продемонстрировали эффективность подходов глубокого обучения. Другие работы [16, 17] использовали рекуррентные нейронные сети с долговременной краткосрочной памятью для различения казахского и русского языков в новостных данных. Их модель, обученная на сегментах длиной всего 2 секунды, достигла точности 86% для казахского языка, что подчеркивает потенциал моделей на основе LSTM в обработке языков с ограниченными ресурсами. Еще одно исследование [18] было сосредоточено на идентификации языка в комментариях казахстанских новостных платформ, где обычно используются казахский и русский языки, часто с переключением кода. Авторы предложили двухэтапную структуру, сочетающую неконтролируемую нормализацию и наивную байесовскую классификацию, а также модель глубокого обучения с использованием сетей LSTM для классификации текста. Их подход улучшил современные результаты по казахскому языку, подчеркнув эффективность сочетания традиционных методов и методов глубокого обучения для многоязычных и смешанных текстов. Эта работа особенно актуальна для продвижения языковой идентификации в сценариях с ограниченными ресурсами и переключением кода. Кроме того, использование i-векторов и x-векторов было широко распространенным подходом в системах идентификации языка и говорящего. Интеграция этих методов была исследована в этой работе [19]. Авторы продемонстрировали надежность систем i-векторов при применении к LID. Несмотря на их успех, ограничения в обработке коротких высказываний и шумной среды привели к дальнейшему развитию в этой области.

Для преодоления проблем языков с низкими ресурсами кросс-языковые модели, такие, как CLSR, стали более эффективным решением. XLSR основан на фреймворке Wav2Vec2, что позволяет совместно использовать речевые представления на нескольких языках, что делает его особенно мощным инструментом для LID в многоязычных условиях [9]. Целью данного исследования является демонстрация улучшений как точности, так и надежности путем тонкой настройки XLSR на различных наборах данных, включая недостаточно представленные языки, такие, как казахский.

3. МЕТОДОЛОГИЯ

В этом разделе рассматриваются наборы данных, используемые для идентификации языка, включая казахский язык. Также в нем рассматриваются методологии, используемые для разработки модели идентификации языка на основе тонкой настройки модели wav2vec2.

В основе модели лежит преобразователь-кодер, который обрабатывает квантованные признаки и фиксирует как локальные, так и глобальные закономерности в речи. Это позволяет модели обрабатывать последовательности различной длины и понимать контекст за пределами отдельных звуков. Кросс-языковая природа XLSR позволяет модели изучать общие представления нескольких языков, что делает ее идеальной архитектурой модели для задач LID. Окончательный вывод создается классификационной головкой, которая определяет наиболее вероятные последующие токены. Эта архитектура позволяет модели достигать высокой точности в идентификации языка, используя общие закономерности в языках с высоким и низким ресурсом.

Набор данных, используемый для этой модели идентификации языка, был составлен из нескольких источников, чтобы обеспечить всестороннее и разнообразное представление языков. Основной набор данных, VoxLingua107 [20], был выбран из-за его обширного охвата многоязычных данных в задачах идентификации языка. Для дальнейшего улучшения набора данных были получены дополнительные данные для русского и английского языков из набора данных Common Voice 17.0 [21]. Для казахского языка использовался Корпус казахской речи 2, разработанный исследовательской лабораторией ISSAI в Назарбаев Университете [22]. Набор данных охватывает шесть языков: казахский, русский, английский, итальянский, испанский и корейский. Эти языки были выбраны на основе их релевантности задаче. Были предприняты согласованные усилия, чтобы гарантировать, что итальянский, испанский и корейский языки

также были представлены в наборе данных для поддержания разнообразия и надежности в задаче идентификации языка.

Для обеспечения высококачественного ввода в модель идентификации языка были предприняты шаги по предобработке данных. Аудиофайлы из каждого набора данных были конвертированы в единый формат WAV с частотой дискретизации 16 кГц, чтобы обеспечить совместимость с моделью wav2vec2. Продолжительность аудиофайлов была приведена к диапазону от 5 до 30 секунд, чтобы избежать перекосов из-за чрезмерно коротких или длинных файлов. Структура набора данных была организована следующим образом: каждая папка соответствовала определенному языку, а внутри нее аудиофайлы и соответствующие текстовые аннотации были структурированы с использованием уникальных идентификаторов для упрощения обработки. Статистические данные по обучающим и тестовым наборам данных приведены в Таблицах 1 и 2.

Таблица 1 – Статистика по набору обучающих данных

Метка	**Всего аудио**	**Общая продолжительность (s)**	**Средняя продолжительность (s)**	**Мин. продолжительность (s)**	**Макс. продолжительность (s)**
kk	42 506	357 510.61	8.41	1.00	24.78
ru	24 969	295 473.74	11.83	1.52	20.00
es	25 439	243 771.40	9.58	1.15	20.00
it	3 181	30 211.77	9.50	2.00	20.00
en	2 677	29 124.43	10.88	1.97	19.98
ko	3 018	27 188.59	9.01	1.72	20.00
Общий	101 790	983 280.54	9.66	1.00	24.78

Таблица 2 – Статистика по тестовому набору данных

Метка	**Всего аудио**	**Общая продолжительность (s)**	**Средняя продолжительность (s)**	**Мин. продолжительность (s)**	**Макс. продолжительность (s)**
kk	10 568	89 112.31	8.43	1.00	24.58
ru	6 354	61 122.64	9.62	1.22	20.00
es	673	7 243.62	10.76	2.37	20.00
it	824	7 964.66	9.67	2.00	19.86
en	6 301	74 041.94	11.75	1.41	20.00
ko	731	6 849.41	9.37	1.81	19.76
Общий	25 448	246 334.57	9.68	1.00	24.58

Набор данных был разделен на обучающий и тестовый наборы, 80% выделено для обучения и 20% зарезервировано для тестирования. Разнообразие и размер набора данных сыграли решающую роль в обеспечении обучения модели на репрезентативном наборе языковых высказываний.

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Обучение проводилось с использованием оптимизатора AdamW со скоростью обучения от 3 до 5. Коэффициент разогрева, равный 0,1, был применен для постепенного увеличения скорости обучения по сравнению с начальными этапами, что важно для достижения стабильного обучения. Модель обучалась в течение 25 периодов времени на одном графическом процессоре, при этом размер пакета для каждого устройства составлял 64, а количество шагов накопления градиента - 4, что эффективно увеличивало размер пакета.

Для оптимизации использования памяти графического процессора и скорости обучения использовалось смешанное прецизионное обучение. Конвергенция обучения и потеря оценки во время обучения модели показаны на Рисунке 1.

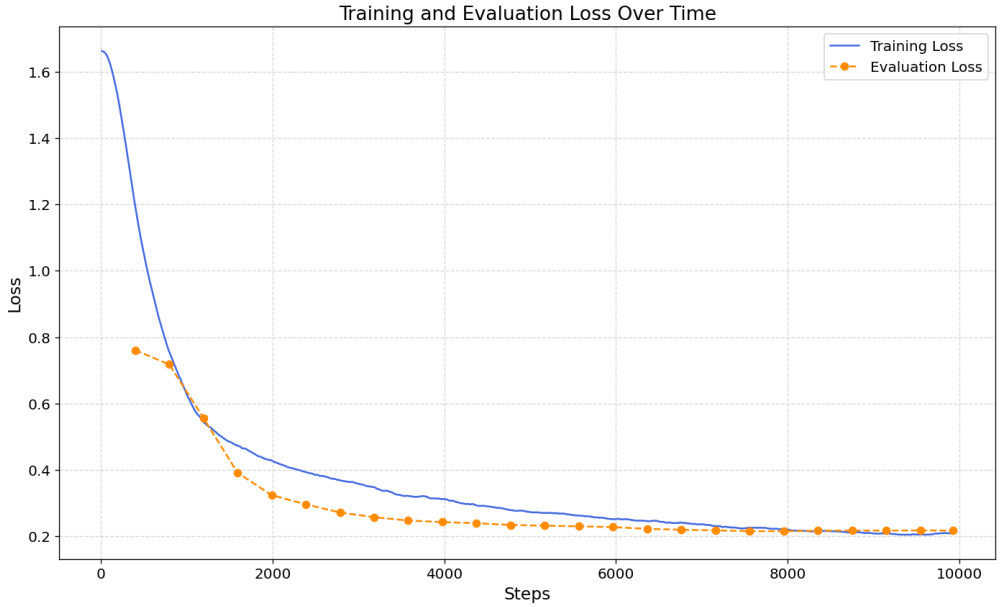

Рисунок 1 – Конвергенция обучения и потеря оценки во время обучения модели

В ходе процесса отслеживались как обучение, так и оценка. Потери обучения неуклонно снижались с течением времени, а потери оценки следовали аналогичной траектории, что указывает на эффективное обучение и хорошую сходимость модели. Ранняя остановка не потребовалась, поскольку не было никаких сигналов переобучения на протяжении 25 эпох. Оценка точно настроенной модели XLSR проводилась путем сравнения ее производительности с несколькими моделями Whisper (Tiny, Small и Large Whisper-3) [23] с использованием того же тестового набора данных. Основными метриками, использованными для сравнения, были точность, прецизионность, отзыв и оценка F1, которые были широко принятыми мерами для задач классификации, таких, как идентификация языка. Сравнение результатов, полученных с помощью моделей, представлено в Таблице 3.

Таблица 3 – Сравнение производительности моделей

Модель	**Точность (Accuracy)**	**Точность (Precision)**	**Полнота (Recall)**	**оценка F1**
Tiny Whisper	0.63	0.90	0.63	0.62
Small Whisper	0.76	0.92	0.76	0.79
Large Whisper	0.85	0.91	0.85	0.86
Fine-tuned Model	**0.93**	**0.93**	**0.93**	**0.93**

Tiny Whisper: несмотря на достижение высокой точности 0,9, она показала относительно низкую точность 0,63 и оценку F1 0,62. Это говорит о том, что, хотя модель могла хорошо предсказывать определенные языки, ее общая способность к полноте и обобщению была ограничена.

Small Whisper: эта модель продемонстрировала разумное улучшение по сравнению с Tiny Whisper с точностью 0,76 и оценкой F1 0,79. Значения полноты и точности были

сбалансированы, что указывает на лучшую производительность в определении более широкого диапазона языков. Однако она все еще отставала с точки зрения общей точности.

Large Whisper-3: как и ожидалось, модель Large Whisper-3 показала значительные улучшения как в точности (0,85), так и в оценке F1 (0,86). Эта модель обеспечила более сильные возможности идентификации языка, но уступила модели XLSR с точки зрения полноты и общей производительности по всем языкам.

Fine-tuned XLSR: Модель XLSR превзошла все другие модели по всем показателям, достигнув точности 0,93 и оценки F1 0,9293. Это указывает на то, что процесс тонкой настройки на разнообразном многоязычном наборе данных, особенно включая языки с низкими ресурсами, такие, как казахский, значительно улучшил способность модели обобщать между языками и обеспечивать высокую производительность в реальных задачах идентификации языка.

Для дальнейшего понимания возможностей модели XLSR был проведен анализ ошибок, чтобы выявить основные причины неверных классификаций. Анализ ошибок модели выявил основные причины неверных классификаций и возможные пути их устранения. Наибольшее количество ошибок возникало при идентификации языков с высокой степенью фонетической и морфологической схожести, таких как испанский и итальянский, шведский и датский, а также русский и казахский, особенно в случаях заимствованных слов. Значительная доля ошибок была связана с низким качеством данных, включая фоновые шумы, обрывки фраз и артефакты в аудиозаписях. Проблемы с классификацией также наблюдались при код-свичинге, когда модель некорректно интерпретировала смешанные фразы, например, на русском и казахском, что снижало точность. Для улучшения качества классификации рекомендуется усилить предварительную обработку данных, включая шумоподавление, добавить больше примеров код-свичинга в тренировочный набор, а также рассмотреть использование дополнительных языковых признаков и механизмы для анализа мультиязычного контекста.

Ключевое преимущество обученной модели XLSR заключается в её способности эффективно работать как с языками с высоким уровнем ресурсов, так и с низкоресурсными языками. В то время как модели, такие, как Whisper, демонстрировали хорошие результаты в условиях высокого уровня ресурсов, использование перекрестного языкового переноса в модели XLSR позволило ей добиться успеха даже с языками, для которых было доступно ограниченное количество данных для обучения, такими, как казахский. Это подчеркивает ценность использования общих представлений между языками, что является неотъемлемой особенностью архитектуры XLSR.

5. ВЫВОДЫ

В этой работе подчеркивается исключительная эффективность тонкой настройки модели XLSR для многоязычной идентификации языков с уделением особого внимания решению проблем, связанных с языками с ограниченными ресурсами. Благодаря стратегическому использованию межъязыкового трансферного обучения модель XLSR значительно превзошла традиционные модели, включая варианты Whisper, достигнув заметных улучшений как в точности, так и в показателях F1. Такая производительность демонстрирует замечательную способность модели справляться с различными лингвистическими сложностями, адаптироваться к изменчивости ресурсов и поддерживать надежность в широком спектре языковых задач. Полученные результаты подчеркивают преобразующую роль XLSR в преодолении разрыва между языками с высокой и низкой ресурсоемкостью, предлагая масштабируемое и эффективное решение для разработки передовых многоязычных систем обработки речи. Эта работа указывает на многообещающий путь к демократизации доступа к передовым языковым технологиям для недопредставленных языков по всему миру.

Вклад авторов

Умбет С. – Исследование, методология, курирование данных, формальный анализ, валидация, визуализация.

Кожирбаев Ж.М. – Концептуализация, курирование данных, формальный анализ, получение финансирования, исследование, методология, ресурсы. Написание - оригинальный черновик, написание - рецензирование и редактирование.

Список литературы

1 Niesler, T. R., Willett, D. Language identification and multilingual speech recognition using discriminatively trained acoustic models // Proceedings of Interspeech. - Pittsburgh, PA, USA, 2006. - P. 134-137.

2 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.

3 Song J., Ermon S. Multi-label contrastive predictive coding // Advances in Neural Information Processing Systems. - 2020. - V. 33. - P. 8161-8173.

4 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.

5 Schneider S., Baevski A., Collobert R., Auli, M. wav2vec: Unsupervised Pre-Training for Speech Recognition // Proceedings of Interspeech. - Graz, Austria, 2019. - P. 3465-3469.

6 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.

7 Fan, Z., Li, M., Zhou, S., Xu, B. Exploring wav2vec 2.0 on Speaker Verification and Language Identification // Proceedings of Interspeech. - Brno, Czechia, 2021. - P. 1509-1513.

8 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.

9 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.

10 Kozhirbayev, Z., Islamgozhayev, T. Cascade speech translation for the Kazakh language // Applied Sciences. - 2023. -V. 13(15). -P. 8900.

11 Kozhirbayev, Z. Kazakh Speech Recognition: Wav2vec2. 0 vs. Whisper // Journal of Advances in Information Technology. - 2023. -V. 14(6). -P. 1382-1389.

12 Kozhirbayev, Z., Karabalayeva, M., Yessenbayev, Z. Spoken term detection for kazakh language // Proceedings of the 4-th International Conference on Computer Processing of Turkic Languages "TurkLang 2016". - Bishkek, 2016. -P. 47.

13 Kozhirbayev, Z., Yessenbayev, Z. Semantically expanded spoken term detection // IEEE Access. -2024. -V. 12. -P. 177844-177855.

14 Singh, G., Sharma, S., Kumar, V., Kaur, B., Bax, M., Masud, M. Spoken Language Identification Using Deep Learning // Computational Intelligence and Neuroscience. - 2021. - V.1. -P. 5123671.

15 Aysa, Z., Ablimit, M., Hamdulla, A. Multi-scale feature learning for language identification of overlapped speech // Applied Sciences. - 2023. - V.13(7). - P. 4235.

16 Kozhirbayev, Z., Yessenbayev, Z., Karabalayeva, M. Kazakh and Russian languages identification using long short-term memory recurrent neural networks // Proceedings of the 11th International Conference on Application of Information and Communication Technologies (AICT). – Moscow, Russia, 2017. -V. 1. -P. 1–5.

17 Kozhirbayev, Z., Yessenbayev, Z., Sharipbay A. Language identification in the spoken term detection system for the kazakh language in a multilinge environment // Journal of Mathematics, Mechanics and Computer Science. - 2019. -V. 96(4). -P. 88–98.

18 Kozhirbayev, Z., Yessenbayev, Z., Makazhanov, A. Document and word-level language identification for noisy user generated text // Proceedings of the 12th International Conference on Application of Information and Communication Technologies (AICT). - Almaty, Kazakhstan, 2018. - P. 1-4.

19 Shen, P., Lu, X., Li, S., Kawai, H. Conditional generative adversarial nets classifier for spoken language identification // Proceedings of Interspeech. - Stockholm, Sweden, 2017. – P. 2814-2818.

20 Valk, J., Alum?e, T. Voxlingua107: a dataset for spoken language recognition // Proceedings of IEEE Spoken Language Technology Workshop (SLT). - Shenzhen, China, 2021. - P. 652-658.

21 Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., Weber, G. Common Voice: A Massively-Multilingual Speech Corpus // Proceedings of the Twelfth Language Resources and Evaluation Conference. -Marseille, France, 2020. - P. 4218-4222.

22 Mussakhojayeva, S., Khassanov, Y., Varol, H. A. KSC2: An industrial-scale open-source Kazakh speech corpus // Proceedings of Interspeech. - Incheon, Korea, 2022. - P. 1367-1371.

23 Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I. Robust speech recognition via large-scale weak supervision // Proceedings of the International conference on machine learning. – Hawaii, USA, 2023. - P. 28492-28518.

Қазақ тіліне арналған Wav2Vec2 моделін пайдалана отырып, ауызша сөйлеу тілін сәйкестендіру

С. Умбет [1] , Ж.М. Кожирбаев [2]
[1] Трир университеті, Университетсринг көшесі 15, 54296, Трир, Германия
[2] National Laboratory Astana, Қабанбай батыр даңғылы, 53, Астана, Қазақстан

Аңдатпа. Бұл зерттеу XLSR (Cross-Lingual Speech Recognition) WAV2VEC2 моделін қолдана отырып, ауызша тілді сәйкестендіру моделін әзірлеуді және дәл баптауды ұсынады. Алты тілді қамтитын бай және алуан түрлі деректер жиынтығында оқытылған, қазақ тілі сияқты ресурстары төмен тілдерге ерекше назар аудара отырып, модель көп тілді сөйлеуді танудың керемет мүмкіндіктерін көрсетеді. Кең бағалаудың арқасында дәл бапталған модель қолданыстағы эталондардан асып қана қоймайды, сонымен қатар басқа заманауи модельдерден, соның ішінде Whisper нұсқаларынан да асып түседі. F1 92,9% және 93% дәлдікпен жоғары нәтижеге қол жеткізген модель өзінің өнімділігін нақты көп тілді және аз ресурстық сценарийлерде көрсетеді. Бұл жұмыс сөйлеуді тану технологияларының дамуына айтарлықтай үлес қосады, әр түрлі тілдік ортада, әсіресе аз ұсынылған тілдік жағдайларда тілді анықтаудың сенімді шешімін ұсынады. Оның жетістігі wav2vec2 негізіндегі модельдердің ресурстары төмен көп тілді контекстерде сөйлеуді өңдеу жүйелерін жақсартудағы әлеуетін көрсетеді. Осы талдаудың қорытындылары қазақ тілі үшін оңтайландырылған сөйлеуді автоматты түрде танудың сенімді және тиімді жүйелерін әзірлеуге ықпал етуі мүмкін. Мұндай технологиялар әртүрлі салаларда, соның ішінде сөйлеуді мәтінге түрлендіруде, дауыстық көмекшілердің жұмысында және дауыстық байланыс құралдарында қолданылады.

Түйін сөздер: тілді сәйкестендіру, ауызша сөйлеу тілін сәйкестендіру, қазақ тілі, Wav2Vec2, XLSR.

Identification of the spoken language using the Wav2Vec2 model for the Kazakh language

S. Umbet [1] , Zh.M. Kozhirbayev [2]
[1] University of Trier, st. Universitetsring 15, 54296, Trier, Germany
[2] National Laboratory Astana, Kabanbay batyr ave. 53, Astana, Kazakhstan

Abstract. This study presents the development and fine-tuning of an oral language identification model using the XLSR (Cross-Lingual Speech Recognition) Wav2Vec2 variant. Trained on a rich and diverse dataset spanning six languages, with a particular focus on low-resource languages such as Kazakh, the model demonstrates remarkable capabilities in multilingual speech recognition. Thanks to extensive evaluation, the finely tuned model not only surpasses existing benchmarks, but also surpasses other modern models, including Whisper variants. Having achieved an impressive F1 score of 92.9% and an accuracy of 93%, the model demonstrates its performance in real multilingual and low-resource scenarios. This work makes a significant contribution to the development of speech recognition technologies by providing a reliable solution for language identification in various language environments, especially in underrepresented language settings. Its success highlights the potential of Wav2Vec2-based models in improving speech processing systems in low-resource multilingual contexts. The results of this analysis can contribute to the development of reliable and effective automatic speech recognition systems optimized for the Kazakh language. Such technologies will find applications in various fields, including speech-to-text conversion, voice assistants and voice communication tools

Keywords: language identification, spoken language identification, Kazakh language, Wav2Vec2, XLSR.

References

1 Niesler T. R., Willett D. Language identification and multilingual speech recognition using discriminatively trained acoustic models, Proceedings of Interspeech, Pittsburgh, PA, USA, 2006. P. 134-137.
2 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.
3 Song J., Ermon S. Multi-label contrastive predictive coding, Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 8161-8173.
4 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.
5 Schneider S., Baevski A., Collobert R., Auli, M. wav2vec: Unsupervised Pre-Training for Speech Recognition, Proceedings of Interspeech, Graz, Austria, 2019. P. 3465-3469.
6 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.
7 Fan Z., Li M., Zhou S., Xu B. Exploring wav2vec 2.0 on Speaker Verification and Language Identification, Proceedings of Interspeech, Brno, Czechia, 2021. P. 1509-1513.
8 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.
9 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.
10 Kozhirbayev Z., Islamgozhayev T. Cascade speech translation for the Kazakh language, Applied Sciences. 2023. Vol. 13(15). P. 8900.
11 Kozhirbayev Z. Kazakh Speech Recognition: Wav2vec2. 0 vs. Whisper, Journal of Advances in Information Technology. 2023. Vol. 14(6). P. 1382-1389.
12 Kozhirbayev Z., Karabalayeva M., Yessenbayev Z. Spoken term detection for kazakh language, Proceedings of the 4-th International Conference on Computer Processing of Turkic Languages "TurkLang 2016", Bishkek, 2016. P. 47.
13 Kozhirbayev Z., Yessenbayev Z. Semantically expanded spoken term detection, IEEE Access. 2024. Vol. 12. P. 177844-177855.
14 Singh G., Sharma S., Kumar V., Kaur B., Bax M., Masud M. Spoken Language Identification Using Deep Learning, Computational Intelligence and Neuroscience. 2021. Vol.1. P. 5123671.
15 Aysa Z., Ablimit M., Hamdulla A. Multi-scale feature learning for language identification of overlapped speech, Applied Sciences. 2023. Vol.13(7). P. 4235.
16 Kozhirbayev Z., Yessenbayev Z., Karabalayeva M. Kazakh and Russian languages identification using long short-term memory recurrent neural networks, Proceedings of the 11th International Conference on Application of Information and Communication Technologies (AICT), Moscow, Russia, 2017. Vol. 1. P. 1–5.
17 Kozhirbayev Z., Yessenbayev Z., Sharipbay A. Language identification in the spoken term detection system for the kazakh language in a multilinge environment, Journal of Mathematics, Mechanics and Computer Science. 2019. Vol. 96(4). P. 88–98.
18 Kozhirbayev Z., Yessenbayev Z., Makazhanov A. Document and word-level language identification for noisy user generated text, Proceedings of the 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, 2018. P. 1-4.
19 Shen P., Lu X., Li S., Kawai H. Conditional generative adversarial nets classifier for spoken language identification, Proceedings of Interspeech, Stockholm, Sweden, 2017. P. 2814-2818.
20 Valk J., Alum?e T. Voxlingua107: a dataset for spoken language recognition, Proceedings of IEEE Spoken Language Technology Workshop (SLT), Shenzhen, China, 2021. P. 652-658.
21 Ardila R., Branson M., Davis K., Henretty M., Kohler M., Meyer J., Morais R., Saunders L., Tyers F. M., Weber G. Common Voice: A Massively-Multilingual Speech Corpus, Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 2020. P. 4218-4222.
22 Mussakhojayeva S., Khassanov Y., Varol H. A. KSC2: An industrial-scale open-source Kazakh speech corpus, Proceedings of Interspeech, Incheon, Korea, 2022. P. 1367-1371.
23 Radford A., Kim J. W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision, Proceedings of the International conference on machine learning, Hawaii, USA, 2023. P. 28492-28518.

Сведения об авторах:

Санжар Умбет – магистрант по специальности "Наука о данных" в Триерском университете, ул. Университетсринг, 15, 54296, Трир, Германия.

Кожирбаев Жанибек Мамбеткаримович – PhD, старший научный сотрудник, National Laboratory Astana, пр. Кабанбай батыра, 53, Астана, Казахстан.

Sanzhar Umbet – masters student in Data Science at University of Trier, st. Universitetsring 15, 54296, Trier, Germany.

Zhanibek Kozhirbayev – – PhD, senior researcher, National Laboratory Astana, Kabanbay batyr ave. 53, Astana, Kazakhstan.

Поступила: 04.12.2024. После редакции: 23.01.2025.
Одобрена: 13.03.2025. Доступна онлайн: 31.03.2025.

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (https://creativecommons.org/licenses/by-nc/4.0/).

Bulletin of L.N. Gumilyov Eurasian National University. Mathematics, computer science, mechanics series, 2025, Vol. 150, №1, P. 17-24.
http://bulmathmc.enu.kz, E-mail: vest_math@enu.kz

Статья
МРНТИ: 27.43.51

СРЕДНИЕ КВАДРАТИЧЕСКИЕ ПОГРЕШНОСТИ ПО МЕРЕ БАНАХА ВОССТАНОВЛЕНИЯ ФУНКЦИЙ КОНЕЧНЫМИ СУММАМИ ЧЛЕНОВ ИХ ТРИГОНОМЕТРИЧЕСКИХ РЯДОВ ФУРЬЕ [1]

Н.Наурызбаев, А.Шоманова * , Н.Темиргалиев
Институт теоретической математики и научных вычислений Евразийского национального университета имени Л.Н. Гумилева, ул. Кажымукана, 13, Астана, 010008, Казахстан
**corresponding author: anarshomanova71@gmail.com,*

Аннотация. В данной статье изучается задача восстановления функций конечными суммами членов их тригонометрических рядов Фурье относительно вероятностных мер на функциональных классах, особенностью которых является невозможность нахождения «*спектра больших коэффициентов*», чем и объясняется рассмотрение произвольных «*конечных сумм из членов рядов Фурье*».

Проблема вероятностного меровведения на классах с индивидуальными оценками на тригонометрические коэффициенты Фурье была решена на основе фундаментального характера свойства "*Функция может быть задана двояко: либо как правило, либо как полный набор тригонометрических коэффициентов Фурье*" из монографии В.М.Тихомирова, остальное было "делом техники".

Переход к последовательностям коэффициентов Фурье с применением теоремы А.Н.Колмогорова о продолжении мер с конечных размерностей пространств на бесконечномерную, позволил ввести вероятностную меру на классах со взвешенными коэффициентами Фурье с доведением до окончательного, впервые построенного Стефаном Банахом в Приложении к книге Станислава Сакса "Теория интеграла". Здесь также предложены некоторые конструктивные детали процесса вероятностного меровведения.

Ключевые слова: тригонометрические ряды Фурье- Лебега, задание функции посредством полного набора его коэффициентов Фурье, вероятностное меровведение на классах функции, востановление функций в среднем квадратическом.

DOI: https://doi.org/10.32523/bulmathenu.2025/1.2

2000 Mathematics Subject Classification: 41A65

1. ВВЕДЕНИЕ

В современной математике существуют два подхода в задачах восстановления: теоретико-функциональный и теоретико-вероятностный. Первый подход заключается, по существу, в нахождении наилучших в том или ином смысле агрегатов приближения, когда за меру действенности вычислительных агрегатов на каком-либо функциональном компакте принимается максимальное уклонение. Однако может получиться так, что «плохо»

[1]Работа выполнена в рамках проекта No. AP19680525, финансируемого Комитетом науки Министерства науки и высшего образования Республики Казахстан

восстанавливаемых функций «мало», но они дают «большую» погрешность – этот эффект на примере поведения положительной непрерывной числовой функции на отрезке ясным образом демонстрирует Теорема о среднем значении функции в Курсе математического анализа. Что в переносе с отрезка на функциональный компакт приводит ко второму подходу, своеобразие которого состоит в том, что сами утверждения об отклонении вычислительных агрегатов от истинного значения приближаемого объекта имеют совершенно иной вид.

Именно, оценивается интеграл по вероятностной мере от погрешности отклонения вычислительных агрегатов от истинного значения приближаемого объекта, в котором в качестве переменной интегрирования выступает функция, пробегающая заданный компакт (см. [1-8] и имеющуюся в них библ.).

Ранее изучалась эффективность квадратурных формул относительно меры Винера и некоторых других мер на классах функций, в контексте сравнения с методом Монте-Карло, здесь же задача восстановления функций.

2. НЕОБХОДИМЫЕ ОПРЕДЕЛЕНИЯ И ВСПОМОГАТЕЛЬНЫЕ УТВЕРЖДЕНИЯ

Пусть даны число $s\,(s=1,2,\dots)$ и положительная последовательность $g\equiv\{g_{m_1,\dots,m_s}\}_{(m_1,\dots,m_s)\in Z^s}$. Определим класс $W\,(g)$ суммируемых и 1-периодических по каждой из s переменных функций $f\,(x)=f\,(x_1,\dots,x_s)$ таких, что

$$\sum_{(m_1,\dots,m_s)\in Z^s}\left|\widehat{f}\,(m_1,\dots,m_s)\right|^2 g^2_{m_1,\dots,m_s}\leq 1, \tag{1}$$

где $(m,x)=x_1m_1+\ldots+x_sm_s$,

$$\widehat{f}\,(m)=\widehat{f}\,(m_1,\dots,m_s):=\int_0^1\dots\int_0^1 f\,(x_1,\dots,x_s)\,e^{-2\pi i(m,x)}dx_1\dots dx_s$$

– тригонометрические коэффициенты Фурье- Лебега.

Сразу же отметим, что в случае сходимости ряда

$$\sum_{m\in Z^s} g_m^{-2}$$

класс $W\,(g)$ состоит из непрерывных функций с абсолютно сходящимися тригонометрическими рядами Фурье

$$\sum_{m\in Z^s}\left|\widehat{f}(m)\right|\leq\left(\sum_{m\in Z^s}\left|\widehat{f}\,(m)\right|^2 g_m^2\right)^{\frac{1}{2}}\left(\sum_{m\in Z^s} g_m^{-2}\right)^{\frac{1}{2}}.$$

Приведем примеры классических классов, содержащихся в $W\,(g)$.

Полагая здесь и всюду ниже $\overline{m}_j=max\,\{|m_j|\,;1\}$, при $g_m=\left(\overline{m}_1^2+\ldots+\overline{m}_s^2\right)^{\frac{r}{2}}$ получаем

Класс Соболева $W_2^r\,(0,1)^s\,(r>0,s=1,2,\dots)$, по определению, есть множество всех 1-периодических по каждой переменной функций $f(x)=f(x_1,\dots,x_s)$ таких, что

$$\sum_{m\in Z^s}\left|\widehat{f}(m)\right|^2\left(\overline{m}_1^2+\ldots+\overline{m}_s^2\right)^r\leq 1.$$

Случай $g_m=(\overline{m}_1\cdot\ldots\cdot\overline{m}_s)^r$ порождает

Класс Соболева с доминирующей смешанной производной $SW_2^r\,(0,1)^s\,(r>0,s=1,2,\dots)$, по определению, есть множество всех 1- периодических по каждой переменной функций $f(x)=f(x_1,\dots,x_s)$, для каждой из которых выполнено неравенство

$$\sum_{m\in Z^s}\left|\widehat{f}(m)\right|^2(\overline{m}_1\cdot\ldots\cdot\overline{m}_s)^{2r}\leq 1.$$

Установим взаимооднозначное соответствие

$$\widehat{f}(m)=\int\limits_{[0,1]^s}f(x)\,cos2\pi\,(m,x)\,dx+i\int\limits_{[0,1]^s}f(x)\,sin2\pi\,(m,x)\,dx\equiv a_m+ib_m\leftrightarrow y_m=(a_m,b_m)\in R^2 \tag{2}$$

между множествами $W(g)$ и при $|y_m|^2=a_m^2+b_m^2$

$$\Omega(g)=\left\{\{y_m\}_{m\in Z^s}:\sum_{m=(m_1,\ldots,m_s)\in Z^s}|y_m|^2\,g_m^2\leq 1\right\}. \tag{3}$$

Определим меру Банаха на множестве s-мерных последовательностей $\Omega(g)$, стало быть, в силу (2), и на функциональном классе $W(g)$.

Пусть $\Gamma\equiv\{\Gamma_k\}_{k=1}^{+\infty}$ есть возрастающая последовательность непустых конечных множеств $\Gamma_k\subset Z^s$, объединение которых есть все Z^s. Через $\nu_k=:|\Gamma_k|$ обозначим число точек в Γ_k.

Пусть $j(m)$ есть последовательное фиксированное упорядочение Γ_1, затем $\Gamma_2\setminus\Gamma_1,\ldots,\Gamma_k\setminus\Gamma_{k-1},\ldots$. Тогда каждый набор ($s$-мерную последовательность)

$$y=\{y_m\}_{m\in Z^s} \tag{4}$$

комплексных чисел, с учетом равенства $y_m=a_{j(m)}+ib_{j(m)}$, будем считать представленной в виде числовой последовательности

$$y=(a_1,b_1,\ldots,a_n,b_n,\ldots). \tag{5}$$

Через $\Omega(g;\Gamma)$ обозначим множество всех наборов (1)-(5) таких, что

$$\sum_{n=1}^{\infty}\left(a_n^2+b_n^2\right)\cdot g_n^2\leq 1,$$

где $\{g_n\}_{n=1}^{+\infty}$ есть подобное (4)- (5) разложение $\{g_{m_1,\ldots,m_s}\}_{(m_1,\ldots,m_s)\in Z^s}$ в числовую одномерную последовательность по $\{\Gamma_k\}_{k=1}^{+\infty}$.

Пусть

$$M_k=\left\{(a_1,b_1,\ldots,a_{\nu_k},b_{\nu_k})\in R^{2\nu_k}:\sum_{n=1}^{\infty}\left(a_n^2+b_n^2\right)\cdot g_n^2\leq 1\right\},$$

а $L(M_k,\mu_k)$ есть σ-алгебра всех измеримых в смысле $2\nu_k$-мерной меры μ_k подмножеств E_k множества $M_k\subset R^{2\nu_k}$.

Измеримое пространство $\Omega(g;\Gamma)$ определим как наименьшую σ-алгебру в $\Omega(g)$, содержащую σ-алгебру $L(M_k,\mu_k)$ цилиндрических множеств $(k=0,1,2,\ldots,E_k\in L(M_k,\mu_k))$

$$S(E_k)=\{(a_1,b_1,\ldots,a_{\nu_k},b_{\nu_k},\ldots)\in\Omega(g):(a_1,b_1,\ldots,a_{\nu_k},b_{\nu_k})\in E_k\}.$$

Согласно Общей теореме Колмогорова о продолжении меры [9],

Каждая из вероятностных мер μ на измеримом пространстве $\Omega(g;\Gamma,\mu)$ однозначно определяется заданием последовательности вероятностных мер μ_k на M_k таких, что для всех $k(k=0,1,2,\ldots)$ и всех $E_k\in L(M_k,\mu_k)$ наряду с равенством

$$\mu\left(S\left(E_k\right)\right)=\mu_k\left(E_k\right)$$

выполнено следующее условие согласованности $(l=1,2,...)$:

$$\mu_k\left(E_k\right)=\mu_{k+l}\left\{\left(a_1,b_1,...,a_{\nu_{k+l}},b_{\nu_{k+l}}\right)\in R^{2\nu_{k+l}}:\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)\in E_k,\sum_{n=0}^{\nu_{k+l}}\left(a_n^2+b_n^2\right)g_n^2\leq 1\right\}. \tag{6}$$

Далее, в случае абсолютно непрерывных, относительно лебеговой, мер μ_k условия (6) приобретают более обозримый вид, позволяющий эффективно строить соответствующие меры.

Именно, в силу теоремы Радона-Никодима для каждого k найдутся определенные на M_k измеримые в смысле $2\nu_k$ -мерной меры Лебега неотрицательные функции $\varphi_k\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)$ такие, что

$$\mu_k\left(E_k\right)=\int_{E_k}\varphi_k, \tag{7}$$

и тогда для μ_k -измеримых множества $T\subset M_k$ и функции $\psi:M_k\mapsto R^1$

$$\int_T\psi\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)d\mu_k\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)=$$
$$=\int_T\psi\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)\varphi_k\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)da_1db_1...da_{\nu_k}db_{\nu_k}. \tag{8}$$

Легко видеть, что тогда условия согласованности (6) перейдут в условия

$$\varphi_k\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)=\int_{A_k}\varphi_{k+1}\left(a_1,b_1,...,a_{\nu_{k+1}},b_{\nu_{k+1}}\right)da_{\nu_k+1}db_{\nu_k+1}\ldots da_{\nu_{k+1}}db_{\nu_{k+1}}, \tag{9}$$

где

$$A_k{=}A_k\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right){=}\left\{\left(a_{\nu_k+1},b_{\nu_k+1},\ldots,a_{\nu_{k+1}},b_{\nu_{k+1}}\right):\sum_{n=\nu_k+1}^{\nu_{k+1}}\left(a_n^2+b_n^2\right)g_n^2\leq 1-\sum_{n=1}^{\nu_k}\left(a_n^2+b_n^2\right)g_n^2\right\}.$$

В свою очередь, функции φ_k, удовлетворяющие условиям (7) и (9), можно задать посредством функций $t_k\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)$ таких, что

$$\int_{A_k}t_{k+1}\left(a_1,b_1,...,a_{\nu_{k+1}},b_{\nu_{k+1}}\right)da_{\nu_k+1}db_{\nu_k+1}\ldots da_{\nu_{k+1}}db_{\nu_{k+1}}=1 \tag{10}$$

для почти всех $\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)\in M_k$, тогда

$$\varphi_k\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)=\prod_{n=1}^{k}t_n\left(a_1,b_1,...,a_{\nu_n},b_{\nu_n}\right), \tag{11}$$

причем выполнение условий (7), (9) эквивалентно выполнению условий (10) -(11).

В итоге, вероятностное меровведение на классе функций на основе понимания функции как полного набора тригонометрических коэффициентов Фурье-Лебега проводится по схеме (1)-(11).

При заданных $\{g_m\}_{m\in Z^s}$, $\Gamma\equiv\{\Gamma_k\}_{k=1}^{+\infty}$ с $\Gamma_k\subset\Gamma_{k+1}\subset Z^s$, мощность $|\Gamma_k|=\nu_k$, $\bigcup_{k=1}^{\infty}\Gamma_k=Z^s$, $\Phi\equiv\{\varphi_k\left(a_1,b_1,...,a_{\nu_k},b_{\nu_k}\right)\}_{k=1}^{+\infty}$ от функции $f\left(x\right)$ как соответствия по правилу f эквивалентным образом переходим к последовательностям $\left\{\widehat{f}\left(m^{(n)}\right)\right\}_{n=1}^{\infty}$ и $\{\left(a_n\left(f\right),b_n\left(f\right)\right)\}_{n=1}^{\infty}$:

$$f(x) \leftrightarrow \left\{\widehat{f}(m)\right\}_{m\in Z^s} \xleftrightarrow{\Gamma\equiv\{\Gamma_k\}_{k=1}^{+\infty}} \left\{\widehat{f}\left(m^{(n)}\right)\right\}_{n=1}^{\infty} \leftrightarrow \left(\widehat{f}\left(m^{(1)}\right),\dots,\widehat{f}\left(m^{(\nu_1)}\right),\right.$$
$$\widehat{f}\left(m^{(\nu_1+1)}\right),\dots,\widehat{f}\left(m^{(\nu_2)}\right),\widehat{f}\left(m^{(\nu_2+1)}\right),\dots,\widehat{f}\left(m^{(\nu_k+1)}\right),\dots,$$
$$\left.\widehat{f}\left(m^{(\nu_{k+1})}\right),\dots\right) \xleftrightarrow{\widehat{f}\left(m^{(n)}\right)=a_n(f)+b_n(f)} \{(a_n(f),b_n(f))\}_{n=1}^{\infty} =$$
$$= (a_1(f),b_1(f),\dots,a_n(f),b_n(f),\dots) = (a_1(f),b_1(f),\dots,a_{\nu_1}(f),b_{\nu_1}(f),\dots,$$
$$a_{\nu_k+1}(f),b_{\nu_k+1}(f),...,a_{\nu_{k+1}}(f),b_{\nu_{k+1}}(f),\dots),$$

далее к интегралам

$$\int\limits_{W(g;\Gamma,\Phi)} \psi(f)\,d\mu_W(f) = \int\limits_{\Omega(g;\Gamma,\Phi)} \psi\left(\widehat{f}\left(m^{(1)}\right),\dots,\widehat{f}\left(m^{(n)}\right),\dots\right) d\mu_\Omega\left(\widehat{f}\left(m^{(1)}\right),\dots,\widehat{f}\left(m^{(n)}\right),\dots\right) =$$
$$= \int\limits_{\substack{\Omega(g;\Gamma,\Phi)=\{(a_1,b_1,\dots,a_n,b_n,\dots):\\ \sum_{n=1}^{\infty}\left(a_n^2+b_n^2\right)\cdot g_n^2\le 1\}}} \psi(a_1,b_1,\dots,a_n,b_n,\dots)\,d\mu_\Omega(a_1,b_1,\dots,a_n,b_n,\dots) =$$
$$= \lim_{k\to+\infty} \int\limits_{\substack{\Omega_k(g;\Gamma,\Phi)=\\ =\{(a_1,b_1,\dots,a_{\nu_k},b_{\nu_k}):\\ \sum_{n=1}^{\infty}\left(a_n^2+b_n^2\right)\cdot g_n^2\le 1\}}} \psi(a_1,b_1,\dots,a_{\nu_k},b_{\nu_k},0,\dots,0,\dots)d\mu_{k,\Omega}(a_1,b_1,\dots,a_{\nu_k},b_{\nu_k},0,\dots,0,\dots) =$$
$$= \lim_{k\to+\infty} \int\limits_{\substack{M_k=\left\{(a_1,b_1,\dots,a_{\nu_k},b_{\nu_k})\in R^{2\nu_k}:\right.\\ \left.\sum_{j=1}^{\nu_k}\left(a_j^2+b_j^2\right)\cdot g_j^2\le 1\right\}}} \psi(a_1,b_1,\dots,a_{\nu_k},b_{\nu_k},0,\dots,0,\dots)\,\varphi_k(a_1,b_1,\dots,a_{\nu_k},b_{\nu_k})$$
$$da_1db_1\dots da_{\nu_k}db_{\nu_k}. \tag{12}$$

Таким образом, задание целого положительного s, s-мерной положительной последовательности $g\equiv\{g_{m_1,\dots,m_s}\}_{(m_1,\dots,m_s)\in Z^s}$, возрастающей к Z^s последовательности непустых конечных множеств $\Gamma\equiv\{\Gamma_k\}_{k=1}^{+\infty}$ и последовательность функций $\Phi\equiv\{\varphi_k\}_{k=1}^{+\infty}$, порождающие согласованные меры $\mu_k(E):=\int_E\varphi_k$ функциональный класс $W(g)$ превращает в пространство с мерой Банаха, которую обозначим через $W(g;\Gamma,\Phi)$.

Отметим, что случай, когда $t_k(a_1,b_1,\dots,a_{\nu_k},b_{\nu_k})$ не зависит от $\left(a_{\nu_{k-1}+1},b_{\nu_{k-1}+1},...,a_{\nu_k},b_{\nu_k}\right)$ изучался С. Банахом [6] (см. также [7-8]).

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Через $\Lambda_k(x;f)=\sum_{m\epsilon\Gamma_k}\widehat{f}(m)\,e^{2\pi i(m,x)}$ обозначим конечную сумму, составленную из членов тригонометрического ряда Фурье-Лебега суммируемой и 1-периодической по каждой из s переменных функции $f(x)$.

Справедлива

Основная теорема. *Предположим, что задано функциональное пространство* $W(g;\Gamma,\Phi)$ *с мерой Банаха* $\mu\equiv\mu_{g,\Gamma,\Phi}=\mu_W$. Тогда для среднего квадратического по мере Банаха $L^2(0,1)^s$ - *уклонения функции* $f(x)$ *от конечной суммы* $\Lambda_k(x;f)$ *членов ее тригонометрического ряда Фурье имеет место равенство* $(g_i=\tilde{g}_{2i-1}=\tilde{g}_{2i})$

$$\int\limits_{W(g;\Gamma,\Phi)} \|f(x)-\Lambda_k(x;f)\|^2_{L^2(0,1)^s}\,d\mu_{g,\Gamma,\Phi}(f) =$$

$$= \sum_{t=k+1}^{\infty} \sum_{\substack{n\,=\,2\nu_{t-1}+1\,= \\ =\,2\,|\Gamma_{t-1}|+1}}^{2\nu_t=2|\Gamma_t|} \int \cdots \int_{A_t} x_n^2 \varphi_t\left(x_1,\ldots,x_{2\nu_t}\right) dx_1 \ldots dx_{2\nu_t},$$

где $A_t = \left\{ (x_1,\ldots,x_{2\nu_t}) : \sum\limits_{i=1}^{2\nu_t} x_i^2 \tilde{g}_i^2 \leq 1 \right\}$.

Доказательство. Применяя теорему Парсеваля, получаем

$$\|f(x) - \Lambda_k(x;f)\|^2_{L^2(0,1)^s} = \left\| \sum_{n\epsilon Z^s \setminus \Gamma_k} \widehat{f}(n)\, e^{2\pi i(n,x)} \right\|^2_{L^2(0,1)^s} = \sum_{n\epsilon\epsilon Z^s \setminus \Gamma_k} \left| \widehat{f}(n) \right|^2.$$

Отсюда

$$\int\limits_{W(g;\Gamma,\Phi)} \|f(x) - \Lambda_k(x;f)\|^2_{L^2(0,1)^s}\, d\mu_W(f) = \int\limits_{W(g;\Gamma,\Phi)} \sum_{n\epsilon\epsilon Z^s \setminus \Gamma_k} \left| \widehat{f}(n) \right|^2 d\mu_W(f) =$$

$$= \int\limits_{W(g;\Gamma,\Phi)} \sum_{t=k+1}^{\infty} \sum_{n\epsilon \Gamma_t \setminus \Gamma_{t-1}} \left| \widehat{f}(n) \right|^2 d\mu_W(f) = \sum_{t=k+1}^{\infty} \int\limits_{W(g;\Gamma,\Phi)} \sum_{n\epsilon \Gamma_t \setminus \Gamma_{t-1}} \left| \widehat{f}(n) \right|^2 d\mu_W(f) =$$

$$\overset{(12)}{=} \sum_{t=k+1}^{\infty} \sum_{\substack{n\,=\,2\nu_{t-1}+1\,= \\ =\,2\,|\Gamma_{t-1}|+1}}^{2\nu_t=2|\Gamma_t|} \int \cdots \int_{\substack{(x_1,\ldots,x_{2\nu_t}) : \\ \sum_{i=1}^{2\nu_t} x_i^2 \tilde{g}_i^2 \leq 1}} x_n^2 \varphi_t\left(x_1,\ldots,x_{2\nu_t}\right) dx_1 \ldots dx_{2\nu_t}.$$

Теорема доказана.

Отсюда, в частности, в случаях $g_m = \left(\overline{m}_1^2 + \ldots + \overline{m}_s^2\right)^{\frac{r}{2}}$ и $g_m = (\overline{m}_1 \cdot \ldots \cdot \overline{m}_s)^r$ следуют соответствуюшие результаты для классов $W_2^r(0,1)^s$ и $SW_2^r(0,1)^s$.

4. ЗАКЛЮЧЕНИЕ

В данной статье представлена схема вероятностного меровведения на классах функций с более подробным изложением в конструктивных деталях, вынесенных в Основную теорему, которая допускает многочисленные конкретизации, включая приведенные здесь классы W и SW.

Вклад авторов: вклад авторов равный.

Список литературы

1 Сульдин А.В. Мера Винера и ее приложения к приближенным методам // Изв. ВУЗов. Математика-I:-1959. -№ 6. -С. 145-158; II: - 1960. -№ 5. -С.165-179.

2 Воронин С.М., Скалыга В.И. О квадратурных формулах // Докл. АН СССР. - 1984. - Т. 246. - № 5. - С. 1038-1041.

3 Smale S. On the efficiency of algorithms of analysis // Bull.Amer.Math.Soc. - 1985. - V.13. - № 2. - P. 87-121.

4 Traub J.F., Wasilkovski G.W, Wozniakovski H. Information-Based Complexity. - New York: Academic Press, 1988.

5 Темиргалиев Н. Концепция С.М.Воронина в проблеме сравнений детерминированных и случайных вычислений в одних и тех же терминах // Вестник Евразийского национального университета имени Л.Н. Гумилева. Серия Математика. Компьютерные науки. Механика. -2019. -Т. 128. № 3. -С. 8-33.

6 Банах С. Интеграл Лебега в абстрактном пространстве// в кн. Сакс С. Теория интеграла. - М.: ИЛ, 1949. - С. 463-477.

7 Воронин С.М., Темиргалиев Н. Об одном приложении меры Банаха к квадратурным формулам // Матем. заметки. - 1986. -Вып. 1. - Т. 39. - С. 52-59.

8 Темиргалиев Н. О построении вероятностных мер на функциональных классах // Труды Матем. ин-та им. В.А. Стеклова РАН. - 1997. - Т. 218. - С. 397-402.

9 Колмогоров А.Н. Основные понятия теории вероятностей. - М.: Наука, 1974.

Функцияларды тригонометриялық Фурье қатарларының мүшелерінің ақырлы қосындылары арқылы жуықтаудың Банах өлшемі бойынша орташа квадраттық қателіктері

Н. Наурызбаев, А. Шоманова, Н. Темірғалиев
Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Теориялық математика және ғылыми есептеулер институты, Қажымұқан көш., 13, Астана, 010008, Қазақстан

Аннотация. Мақалада функцияларды олардың тригонометриялық Фурье қатарларының ақырлы қосындылары арқылы «*үлкен коэффициенттер спектрін*» анықтау мүмкін еместігінен Фурье қатарларының кез келген «*ақырлы қосындыларын*» зерттеу қажеттілігімен ерекшеленетін функциялар класстарында анықталған ықтималдық өлшемдер бойынша жуықтау мәселесі қарастырылады.

Тригонометриялық Фурье коэффициенттеріне жеке бағалар берілген функциялар класстарында ықтималдық өлшемдерді енгізу мәселесі В.М. Тихомировтың монографиясындағы «*Функциялар екі түрлі тәсілмен беріледі: бірі – ереже түрінде болса, екіншісі – тригонометриялық Фурье коэффициенттерінің толық жиыны арқылы*» деген қағидасын негізге алады. Қалғаны – "техникалық мәселе" ғана.

Фурье коэффициенттерінің тізбектеріне өту және А.Н. Колмогоровтың өлшемдерді ақырлы өлшемді кеңістіктерден ақырсыз өлшемді кеңістікке дейін жалғастыру туралы теоремасын қолдану арқылы Фурье коэффициенттері салмақталған класстарда ықтималдық өлшем енгізуге мүмкіндік берілді, оны алғаш рет Стефан Банах Станислав Сакстың "Интегралдар теориясы" кітабындағы қосымшада құрды. Сонымен қатар, мақалада ықтималдық өлшемді енгізу процесінің кейбір конструктивті қырлары да ұсынылған.

Түйінді сөздер: Фурье- Лебег тригонометриялық қатарлары, функцияны оның Фурье коэффициенттерінің толық жиыны арқылы беру, функциялар класстарында ықтималдық өлшемін енгізу, орташа квадраттық мағынада функцияны жуықтау.

Average Square Errors by Banach Measure of Recovery of Functions by Finite Sums of Terms of Their Trigonometric Fourier Series

N. Nauryzbaev, A. Shomanova, N. Temirgaliyev
Institute of Theoretical Mathematics and Scientific Computations, L.N. Gumilyov Eurasian National University, Kazhymukan str., 13, Astana, 010008, Kazakhstan

Abstract. The paper considers the problem of recovering functions using finite sums of terms of their trigonometric Fourier series with respect to probability measures on function classes characterized by the inability to determine a "*spectrum of large coefficients*". This difficulty motivates the consideration of arbitrary "*finite sums of terms from the Fourier series*".

The problem of defining probability measures on classes with individual estimates for trigonometric Fourier coefficients was resolved based on the fundamental nature of the property "*A function can be defined in two ways: either by a rule or by the complete set of its trigonometric Fourier coefficients*", as discussed in the monograph by V.M. Tikhomirov, The remaining steps were largely a matter of technical execution.

The transition to sequences of Fourier coefficients, using A.N. Kolmogorov's extension theorem (on the extension of measures from finite-dimensional to infinite-dimensional spaces), allowed the introduction of a probability measure on classes with weighted Fourier coefficients, ultimately leading to the Banach measure – first constructed by Stefan Banach in the appendix to Stanisław Saks' book "Theory of the Integral". This paper also proposes some constructive details of the process of probabilistic measure

Keywords: Fourier-Lebesgue trigonometric series, defining a function employing a complete set of its Fourier coefficients, probabilistic measurement on function classes, recovery of functions in average square.

Список литературы

1 Suldin A.V. Mera Vinera i eye prilozheniya k priblizhennym metodam [Wiener's measure and its appli- cations to approximate methods], Izv. VUZov[Izv. Universities]. Matematika-I: (6), 145-158(1959); II: (5), 165-179(1960).

2 Voronin S.M.. Skalyga V.I. O kvadraturnykh formulakh [On quadrature formulas], Dokl. AN SSSR [Dokl. USSR Academy of Sciences], 246(5), 1038-1041(1984).

3 Smale S. On the efficiency of algorithms of analysis, Bull. Amer. Math. Soc., 13(2), 87-121(1985).

4 Traub J.F., Wasilkovski G.W, Wozniakovski H. Information-Based Complexity. (Academic Press, New York, 1988).

5 Temirgaliyev N. The concept of S.M.Voronin in the problem of comparisons of deterministic and random computation in the same terms, Bulletin of the L.N. Gumilyov Eurasian National University.Mathematics. Computer Science. Mechanics Series. 2019. Vol. 128. № 3. P. 8-33.

6 Banakh S. Integral Lebega v abstraktnom prostranstve [The Lebesgue integral in abstract space]// v kn. Saks S. Teoriya integrala [in the book. Saks S. Theory of the integral]. Moscow, IL. 1949. P. 463-477

7 Voronin S.M., Temirgaliev N. Application of Banach measure to quadrature formulas, Mathematical notes of the Academy of Sciences of the USSR, 39(1), 30-34(1986).

8 Temirgaliyev N. O postroyenii veroyatnostnykh mer na funktsionalnykh klassakh[On the construction of probability measures on functional classes], Trudy Matem. in-ta im. V.A. Steklova RAN [Proceedings of a mathematical institute V.A. Steklov RAS.],218, 397-402(1997).

9 Kolmogorov A.N. Basic concepts of probability theory [Osnovnyye ponyatiya teorii veroyatnostey]. (Nauka, Moscow, 1974).

Сведения об авторах:

Наурызбаев Нурлан Жумабаевич – PhD, Старший научный сотрудник Института теоретической математики и научных вычислений Евразийского национального университета им. Л.Н. Гумилева, Кажымукана, 13, Астана, 010008, Казахстан.

Шоманова Анар Аманаевна – Автор для корреспонденции, Старший научный сотрудник Института теоретической математики и научных вычислений Евразийского национального университета им. Л.Н. Гумилева, Кажымукана, 13, Астана, 010008, Казахстан.

Темиргалиев Нурлан – доктор физико-математических наук, профессор, директор Института теоретической математики и научных вычислений Евразийского национального университета им. Л.Н. Гумилева, Кажымукана, 13, Астана, 010008,Казахстан

Nauryzbaev Nurlan – PhD, Senior Researcher, Institute of Theoretical Mathematics and Scientific Computations, L.N. Gumilyov Eurasian National University, Kazhymukan str., 13, Astana, 010008, Kazakhstan.

Shomanova Anar – Corresponding author, Senior Researcher, Institute of Theoretical Mathematics and Scientific Computations, L.N. Gumilyov Eurasian National University, Kazhymukan str., 13, Astana, 010008, Kazakhstan.

Temirgaliyev Nurlan – Prof., Doctor of Phys. -Math. Sciences, Director of the Institute of Theoretical Mathematics and Scientific Computations, L.N. Gumilyov Eurasian National University, Kazhymukan str., 13, Astana, 010008, Kazakhstan.

Поступила: 13.01.2025. После редакции: 23.02.2025.
Одобрена: 13.03.2025. Доступна онлайн: 31.03.2025.

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (https://creativecommons.org/licenses/by-nc/4.0/).

Bulletin of L.N. Gumilyov Eurasian National University. Mathematics, computer science, mechanics series, 2025, Vol. 150, №1, P. 25-41.
http://bulmathmc.enu.kz, E-mail: vest_math@enu.kz

Article
IRSTI: 27.43.51

Mathematical methods as a basis and scientific justification for the development of an information-analytical database for the accounting and analysis of OSH expenditures [1]

Abikenova Sh. [1], Aitimova Sh. [2], Kocherbaeva A. [3]
[1] *«Scientific Research Company GALYM» LLP, G. Musrepov str., 15a, Astana, Republic of Kazakhstan*
[2] *JSC «State Social Insurance Fund», Dostyk str., 18, Astana, Republic of Kazakhstan*
[3] *Kyrgyz-Russian Slavic University named after B.N. Yeltsin, Kievskaya str., 44, Bishkek, Kyrgyz Republic*
nikgalym@gmail.com

Abstract. In contemporary conditions, there is a growing interest in the study of economic losses resulting from occupational injuries and diseases across various sectors of the economy. Mathematical methods serve as the primary research tools, and they are applied in calculations, structural and dynamic analyses, and related areas. The economic impact of such losses, along with their inverse relationship to investments in occupational safety (i.e., preventive expenditures), plays a crucial role in preserving workers' health, enhancing labor productivity, and fostering overall economic growth. This substantiates the need for continuous improvement in occupational safety and health (OSH) through various instruments, one of which is the comparative analysis of enterprise OSH expenditures, particularly those aimed at improving working conditions over a given period.

The article substantiates the development of an information-analytical database for the accounting and analysis of OSH expenditures based on the application of mathematical methods, including comparison, classification, and systematization, through the juxtaposition of primary data on key OSH expenditure items. The categories of expenditures analyzed are defined in accordance with the labor legislation of the Republic of Kazakhstan. The purpose of the study is to apply mathematical methods in the development of a software solution for OSH cost accounting and analysis.

Keywords: mathematical methods, information-analytical database, occupational safety and health cost accounting and analysis, baseline indicators, software solution.

DOI: https://doi.org/10.32523/bulmathenu.2025/1.3

2000 Mathematics Subject Classification: 93C42

1. Introduction

In Kazakhstan, a national occupational safety and health (OSH) policy and program have been established in accordance with the provisions of ILO Convention No. 187 entitled «Promotional Framework for Occupational Safety and Health». This commitment is reflected in the Concept for Safe Labor until 2030, approved by the Resolution of the Government of the Republic of Kazakhstan dated December 28, 2023, No. 1182 [1, 2]. The Concept outlines the main directions of action to be taken at various levels of governance. A key distinguishing feature of the newly adopted national

[1]The scientific results were obtained within the framework of grant funding for a research project by the Ministry of Science and Higher Education of the Republic of Kazakhstan entitled «Improving the System of Accounting and Analysis of Occupational Safety Expenditures in High-Risk Industries» (Project No. AP19680581).

OSH program is that, for the first time at the national level, it emphasizes the need to introduce new financial and economic measures aimed at increasing the employer's contribution to improving working conditions. According to paragraph 9 of the Action Plan for the implementation of the Concept for Safe Labor in Kazakhstan for 2024–2030, it is envisaged to introduce approaches to economic planning of occupational safety and health expenditures.

It is important to note at the outset that a scholarly review of international approaches to the accounting and analysis of OSH expenditures has revealed that the terms «costs» and «expenses» carry distinct economic meanings [3]. The term «costs» refer to the value of various types of resources (material, financial, informational, labor, and others) utilized for production or the maintenance of operations. In contrast, expenses are those costs that have been paid and can be both documented and economically justified by the enterprise. In the context of OSH, this article will henceforth use these distinctions as a basis for further discussion.

The provisions of the Labour Code of the Republic of Kazakhstan dated November 23, 2015, No. 414-V ZRK [4], specifically Article 182, impose a number of obligations on employers. These include conducting an assessment of occupational risks and implementing measures to minimize or eliminate them through prevention, replacement of production equipment and technological processes with safer alternatives, providing training, briefings, and knowledge testing for workers on OSH, and ensuring the availability of documentation for safe conduct of production processes and operations.

Employers are also required to organize OSH training and knowledge assessments for managers and personnel responsible for OSH, at least once every three years; to provide workers with adequate sanitary and hygienic conditions, supply and repair of personal protective clothing and footwear, prophylactic treatments, cleansing and disinfecting agents, medical first-aid kits, milk or equivalent nutritional products, and/or specialized dietary (therapeutic and preventive) foods, as well as personal and collective protective equipment in accordance with standards established by the authorized state labor body.

Furthermore, employers must record, monitor, and analyze occupational accidents and diseases (OAD); conduct, with the participation of workers' representatives, periodic certification of workplaces with regard to working conditions at least once every five years, following regulations approved by the authorized labor authority; insure employees against occupational accidents and report the conclusion of such mandatory insurance contracts to the authorized labor authority under established procedures.

Additional responsibilities include the development, approval, and revision of OSH instructions in accordance with procedures set by the authorized labor authority; financing and conducting mandatory and periodic (throughout the course of employment) medical examinations, as well as pre-shift health screenings where required by agreements, collective contracts, or national legislation, or when transferring to work under different conditions or upon the emergence of signs of occupational illness; and implementing and monitoring the functioning of an OSH management system.

The source of funding for this broad range of employer obligations related to ensuring safe working conditions is the employer's own financial resources. Given the extensive scope of these OSH-related obligations, which require adequate funding, there is a pressing need for regular economic assessments and comparative analyses.

In this context, the article explores various applied mathematical methods for the accounting and analysis of the aforementioned OSH expenditures. It also provides a scientific rationale for the development of an information-analytical database for tracking OSH expenditures, the modification of reporting practices related to OSH expenditure accounting and analysis, and the development of a mathematical algorithm for accounting for costs associated with occupational accidents and for processing data within the information-analytical system.

2. Methods

Given the diversity of approaches to classifying OSH expenditures, it is recommended that cost accounting be based on the direct attribution of expenditures to OSH activities and the specifics of budgeting for both mandatory and special-purpose measures (the latter being applicable only

in the presence of hazardous or harmful working conditions as identified through occupational risk assessment). These expenditures may be grouped into three categories:

1. General OSH expenditures;
2. Special OSH expenditures;
3. Economic losses related to OSH.

General OSH expenditures are mandatory for all enterprises.

Special OSH expenditures are compulsory when hazardous and/or harmful working conditions are present in the enterprise.

Economic losses in the OSH domain include costs associated with penalties for non-compliance with OSH legislation and compensation payments to workers affected by occupational accidents.

Thus, based on the review, two main approaches have been identified:

Preventive expenditures («investment in OSH»), referring to costs aimed at ensuring safe working conditions and implementing OSH-related activities. These typically include expenses such as training, insurance, personal and collective protective equipment, occupational risk assessments, periodic medical examinations, and pre-shift health screenings, as well as the provision of milk or equivalent food products and/or specialized dietary (therapeutic and preventive) nutrition. These are expenditures intended to prevent the occurrence of a occupational accident.

Compensatory expenditures («economic losses»), referring to costs incurred as a result of an actual occupational accident. These generally include compensation payments, penalties, and related expenses. Concerning the accident, these are post-factum costs that carry a negative economic character.

The methodological foundation of this study is based on the cost analysis framework for OSH in industries with a high risk of occupational injuries [5]. The analysis is conducted using comparative methods, with baseline indicators for the reporting period (a minimum of two years, with a recommended three-year dynamic analysis) serving as the basis for calculations and assessments.

OSH cost analysis may be carried out using either a horizontal or vertical (structural) format, and may also involve the use of ratio analysis.

In horizontal analysis, both absolute and relative deviations across cost items are examined. This approach allows for the comparison of planned versus actual expenditures, previous periods versus the current reporting period, or all of these combined, thereby enabling a comprehensive evaluation of cost dynamics.

In vertical (structural) analysis, the obtained data are compared across individual cost items, allowing for cross-comparison between various categories from all types of reports. This type of analysis helps identify the factors contributing to increases or decreases in expenditures.

During ratio analysis, specific indicators are calculated, the values of which can be compared over different periods. These ratios can also be benchmarked against national or sector-specific averages.

The following formulas (1)–(3) may be applied for performing the necessary calculations:

$$Pb = \frac{Zf}{Zb} * 100\%, \tag{1}$$

where

Pb – percentage relative to the baseline value,

Zf – actual value,

Zb – baseline value.

According to the formula, the percentage relative to the baseline value is calculated as the ratio of each item to the selected baseline value.

The implementation of horizontal analysis involves comparing indicators in both absolute and relative terms.

At the initial stage of the analysis, absolute deviations of financial reporting items are calculated relative to the baseline period.

$$\Delta abs = Zf - Zb, \tag{2}$$

where

Δ abs – absolute change.

At the second stage, relative changes in the analyzed indicators are calculated in comparison with the baseline period.

$$\Delta rel = \frac{\Delta abs}{Zb} * 100\% \tag{3}$$

In this case, the absolute deviations are normalized relative to the baseline period value.

This approach is particularly useful when comparing multiple companies that differ in size and scale of operations. The total amount of OSH expenditures $Z_i(i = 1, \ldots)$ for the i-th enterprise is determined using formula (4).

$$Z_i = \sum_{k=1}^{n} S_k + \sum_{j=1}^{m} K_j, \tag{4}$$

where

S_k (k=1,..n) – preventive expenditures, n – number of expenditure types;

K_j (j=1,. . . m) - compensatory expenditures, m – number of expenditure types.

3. Results and Discussion

The modern enterprise accounting system comprises three primary types: financial accounting, managerial accounting, and tax accounting. It is from this perspective that a methodological overview of approaches to OSH cost accounting is presented.

Financial accounting is employed to prepare information on the enterprise's financial and economic activities for both internal and external use. It is mandatory and strictly regulated by the state [6], as reporting is submitted in accordance with officially established formats.

Tax accounting is essential for the accurate determination of the tax base and the calculation of mandatory tax payments. This type of accounting is also mandatory and governed by legislation, specifically the Tax Code of the Republic of Kazakhstan [7]. The rules for maintaining records and submitting reports are defined at the state level because tax accounting ensures accurate taxation of each organization and enables control over tax remittances. Thus, the primary objective of tax accounting is to accurately reflect cash flows to determine the amount of taxes owed and report this information to the state in a prescribed format.

Managerial accounting refers to the collection and use of information required by a company's management to support strategic decision-making processes.

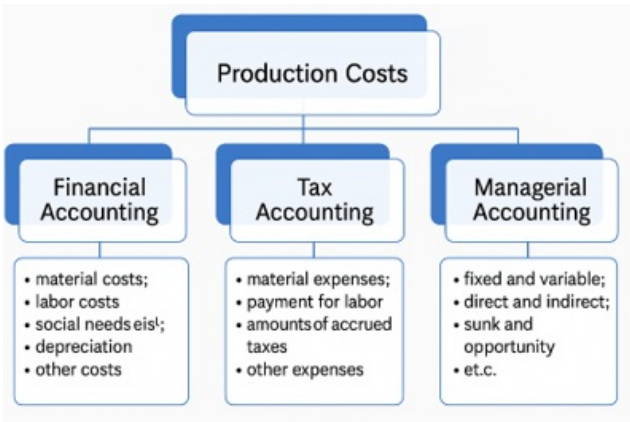

Figure 1 – Types of Cost Accounting

Taking into account the key distinctions among the three types of accounting within organizations—financial, tax, and managerial accounting—each is defined by its specific objectives. Financial accounting is maintained to record all financial transactions of the company to produce

accurate and reliable financial statements. The purpose of tax accounting is to optimize the company's tax liabilities and ensure compliance with tax legislation. Managerial accounting is organized to provide information for managerial decision-making and internal control.

In this context, modern organizations are increasingly implementing the integration of financial, tax, and managerial accounting systems, enabling a comprehensive approach to enterprise management.

Firstly, a unified data platform allows all three types of accounting to operate within a common information system, providing real-time access to relevant data.

Secondly, cross-functional reporting enables the use of data from different accounting systems to generate integrated reports. This facilitates more informed decision-making by management, taking into account financial, tax, and managerial perspectives.

Thirdly, process synergy arises when the optimization of one type of accounting (e.g., tax accounting) positively impacts another (e.g., managerial accounting), as cost and revenue information can be used both to reduce tax liabilities and to improve operational efficiency.

Fourthly, integration enhances analytical and forecasting capabilities, allowing for a more holistic assessment of a company's financial activities.

P. Richardsson [8] emphasizes that effective OSH cost accounting requires a comprehensive approach and that investments in OSH lead to a reduction in workplace accidents, decreased downtime, and increased productivity, ultimately generating substantial long-term economic benefits.

Cameron A. Mustard and B. Yanar [9] apply a transparent methodology to assess the financial benefits of OSH expenditures, highlighting that even moderate investments in OSH can yield significant economic advantages for employers.

Raini D. Wirahadikusumah and Felix Adhiwirya [10] examine OSH-related costs associated with the implementation of safety management systems in construction projects. The authors note that mandatory OSH management system costs can range from 2.01% to 3.70% of the contract value, with the most significant components being expenditures on OSH personnel, insurance, and licensing, as well as training and the promotion of OSH programs.

Particular interest is drawn to the work of V.L. Pavlova [11], which demonstrates the correlation between OSH expenditure levels and the degree of occupational risk. To determine the required volume of financial resources, the author proposes a formula for calculating costs by classification group, based on the existing and projected levels of occupational risk. The level of occupational risk is treated as a variable necessary to link OSH expenditures to the actual performance outcomes of the enterprise in the field of OSH.

The article further presents an analysis of costs associated with OAD, which entail severe losses, including direct physical harm to the injured individual, emotional distress for their family and community, and damage to societal values such as justice and solidarity. While attempts are made to assign monetary value to these losses (a necessity, especially in legal proceedings aimed at determining damages), in the end, no financial figure can truly capture the irreparable harm that cannot be restored through compensation [12].

To assess the magnitude of economic losses and analyze their impact on the enterprise and related organizations, various methods and models are applied in global practice. These tools provide insight into losses resulting from worker absenteeism, occupational accidents, staff turnover, loss of work capacity, changes in working conditions, and other factors. The same methods and models can also be employed to analyze the costs of OSH improvement measures and to determine the economic efficiency of such interventions.

A substantial body of scientific literature is devoted to methodologies for estimating the cost of OAD. The estimation of these costs was the focus of research by H. Heinrich [13], who demonstrated that the costs associated with workplace accidents in American companies were significant, and that many of these expenses remained hidden from management.

Numerous similar studies have been conducted following H. Heinrich's work in various countries—for example, by Grimaldi and Simonds [14], Larsson and Betts [15], and N. Monnerie [16]. These studies document that substantial costs may be associated with OAD.

A frequently cited rationale for measuring OSH-related expenditures is that making such costs explicit encourages managers to consider OSH issues in their decision-making processes. In other words, OSH and workplace safety concerns become more closely aligned with core business operations and exert a greater influence on managerial decisions.

Overall, in the literature on this topic, OSH-related costs are generally divided into two broad categories:

Costs associated with the functioning of the OSH management system and initiatives aimed at promoting and sustaining a culture of workplace safety within the company. These costs are typically relatively stable, as they are not directly affected by the occurrence of adverse events such as OAD.

Costs resulting from OAD. These may take the form of direct expenditures, increased operational costs, potential revenue losses, or opportunity costs. Such costs vary depending on the type and number of adverse outcomes.

This categorization is reflected in a range of studies dedicated to the assessment of OSH-related expenditures.

D. Andreoni [17], along with V. Khrimak and H. Perezgonzalez [18], employ the terms «prevention costs» and «accident costs», where prevention costs are not contingent on the occurrence of individual accidents. This terminology is, to some extent, comparable to the cost-of-quality framework used by B. Dale and J. Plunkett [19].

However, it should be noted that certain expenditures may fall into both categories. For instance, some insurance-related costs include fixed annual premiums as well as variable components that increase with the frequency or severity of claims.

According to the International Financial Reporting Standards (IFRS), an entity shall present an analysis of expenses recognized in profit or loss using a classification based either on their nature or their function within the entity, depending on which method provides information that is reliable and more relevant.

An entity must present an analysis of expenses recognized in profit or loss using a classification based either on the nature of the expenses or on their function, depending on which approach ensures more reliable and relevant information.

The nature of the expense method involves grouping expenses recognized in profit or loss according to their nature (e.g., procurement of personal protective equipment (PPE), social security costs) without allocating them to the functional areas of the entity.

This method may be simpler to apply, as it does not require the reallocation of expenses based on their purpose or function within the organization.

The function of the expense method classifies expenses according to their purpose. An entity that classifies expenses by function is required to disclose additional information on the nature of the expenses, including those related to the acquisition or use of PPE.

These methods provide insights into costs that may vary, either directly or indirectly, with changes in the enterprise's sales or production volume. Since each presentation method has its own advantages depending on the type of entity, the standard requires management to choose the presentation format that is both reliable and most relevant.

Expense analysis is carried out based on an analytical model consisting of 17 cost elements, aligned with the labor legislation of the Republic of Kazakhstan regarding the provision and financing of OSH measures.

The analysis covers a range of expenses, including:

1. costs associated with the assessment of occupational risks;
2. certification of production facilities with respect to working conditions;
3. training, instruction, and knowledge testing on OSH issues for workers, managers, and personnel responsible for OSH;
4. mandatory insurance of employees against accidents occurring in the course of performing their professional (official) duties (MAI);
5. provision of personal protective equipment (PPE);
6. provision of collective protective equipment;
7. provision of sanitary and welfare facilities and installations;

8. distribution of milk or equivalent food products and/or specialized dietary (therapeutic and preventive) products to employees;
9. granting of reduced working hours, additional paid annual leave, and increased wage rates;
10. periodic medical examinations and pre-shift health screenings;
11. payment of mandatory professional pension contributions.

This section briefly presents the practical results of testing the proposed methodology. A total of 240 distinct OSH-related data points were processed; dynamic trends over three years were analyzed across 119 indicators; and econometric indicators were calculated based on 167 data entries. All numerical characteristics were further subjected to graphical analysis. As a result, an information-analytical database for the accounting and analysis of OSH expenditures was developed (Table 1).

Based on the information-analytical database for the accounting and analysis of OSH expenditures, a cost analysis was conducted in accordance with the established methodology.

The OSH expenditures analysis for **Enterprise A**—a large-scale enterprise employing a total of 4,826 workers, of whom 4,714 (or 97%) are engaged in hazardous working conditions—revealed that, during the analyzed period, expenditures were allocated across 14 out of 17 identified cost categories. On average over the three years, the dominant OSH expenditure items were: provision of PPE - KZT 276.32 million (29%), MPPC - KZT 252.93 million (26%), and provision of additional paid leave -KZT 252.90 million (26%).

The nature of OSH expenditures reflects the working conditions at the enterprise, which are classified as hazardous and/or harmful.

Total expenditures on benefits and compensation for employees working under hazardous conditions at the enterprise account for 61% of overall OSH spending, which is attributable, in part, to the increase in the number of employees identified as working in hazardous environments based on the results of workplace certification, from 4,299 to 4,686 individuals.

A significant rise was observed in the costs of advanced training for managers and OSH-responsible personnel, due to an increase in the number of trainees from 277 to 357 persons.

Expenditures on MAI have been increasing annually by approximately 20%, reaching a total of KZT 85.1 million in 2019. These insurance costs are calculated as 0.76% of the total wage fund, based on the classification of the enterprise's economic activity under Risk Class 12, which determines the premium under the insurance contract. Positive growth trends were also recorded for expenditures on: PPE, cleansing and neutralizing agents (24%), sanitary and welfare services (25%), and periodic medical examinations (12%).

The OSH expenditure analysis for **Enterprise B**, which employs a total of 242 workers—41 of whom (16%) are engaged in hazardous working conditions—shows that, during the analyzed period, funds were allocated to 11 out of 17 categories of OSH expenditures (see Table 1).

Over the three years, the dominant OSH expenditure items at Enterprise B were: provision of PPE – KZT 8.75 million (43%), pre-shift medical screenings – KZT 2.59 million (13%), payment of MPPC – KZT 2.45 million (12%). The expenditures on PPE were related to the procurement of 368 types of special clothing, protective footwear, and other PPE. A comparative analysis of OSH expenditure dynamics reveals a baseline growth rate of 26.2%. The increase is primarily driven by rising costs in the following areas: MAI – up by 129.6% (under the mandatory insurance contract, the insurance premium rate is 0.65%, with the enterprise classified under Professional Risk Class 8); granting of additional paid leave for employees working in hazardous and/or dangerous conditions – 69.6% increase; MPPC – 31.1% increase.

These trends are linked to the expansion of the workforce, following the commissioning of an additional production unit, and the increase in the total wage fund due to higher wages. This growth in wages has directly impacted the volume of expenditures on MPPC, MAI, additional leave entitlements, and increased wage payments.

A violation of labor legislation was noted with regard to internal OSH oversight. Specifically, at this industrial enterprise employing more than 50 people, an OSH service has not been established, and OSH responsibilities are being carried out by an employee assigned on a part-time basis.

Table 1 – Information-analytical database for the accounting and analysis of OSH expenditures at six enterprises (A–G) over a five-year period

Names of Cost Items	Enterprises													
	A		B		C		D		E		F		G	
	Exp., KZT*	Share, %**	Exp., KZT*	Share, %**	Exp., KZT*	Share, %**	Exp., KZT*	Share, %**	Exp., KZT*	Share, %**	Exp., KZT*	Share, %**	Exp., KZT*	Share, %**
Provision of PPE, cleansing, and neutralizing agents	276 323 539	29	8 751 667	43	1 074 333	4	1 063 804 495	16	22 468 819	8	1732600	42	31876694	20
Provision of sanitary and welfare services	1 948 948	0	268 333	1	335 333	1	425 672 333	6	247 168	0	820000	20	746539	0
Certification of production facilities based on working conditions	5 624 788	1	1 824 095	9	179 667	1	6 672 000	0	1 408 693	0	0	0	1111607	1
Occupational risk assessment				0		0		0		0	0	0		0
Periodic medical examinations	5 052 013	1	870 332	4	802 000	3	98 692 333	1	4 723 619	2	75000	2	9850140	6
Pre- and post-shift medical screenings	2 900 000	0	2 593 000	13	65 667	0	41 177 000	1		0	0	0		0
Mandatory accident insurance	74 323 971	8	1 674 733	8	3 212 333	13	154 554 333	2	38 064 688	13	864792	21	9034122	6
Advanced OSH training	1 413 667	0	324 000	2	312 667	1	6 368 000	0	500 000	0	90000	2		0
Training, instruction, and knowledge testing				0		0	15 352 333	0	10 081 667	3	59000	1	1586567	1
Equipping OSH and safety training rooms	1 056 631	0		0	145 667	1	6 784 000	0		0	351000	9		0
Distribution of milk or equivalent food products	62 475 467	7	638 333	3	2 131 667	8	3 065 333	0	11 641 663	4	130000	3	27813139	18
Provision of therapeutic and preventive nutrition (TPN) and vitamin supplements	2 408 000	0		0		0	544 776 333	8		0		0		0
Granting of additional paid leave	252 901 600	26	1 029 333	5	15 863 400	62	345 792 333	5	59 627 513	20		0	15270550	10
Increased wage rates				0		0	3 102 026 333	46	106 075 217	36		0	41949300	27
Reduced working hours	15 559 576	2				0	517 374 000	8		0		0		0
Payment of mandatory professional pension contributions (MPPC)	252 931 802	26	2 446 000	12	1 556 333	6	354 338 667	5	37 421 474	13		0	17880984	11
Consulting services	7 280 361	1				0		0		0		0	31876694	20
Total	959 773 575	100	20 419 826	100	25 679 807	100	6 686 449 829	100	292 260 520	100	4 122 392	100	157 119 642	100

* averaged over a five-year period

** share in total expenditures, averaged over a five-year period for each enterprise

The OSH expenditure analysis for **Enterprise C**, which employs a total of 236 workers – 87 of whom (36%) are engaged in hazardous working conditions – shows that over the three years, the dominant OSH expenditure items were: provision of additional paid leave for employees engaged in heavy work or work under hazardous and/or dangerous conditions – KZT 15.86 million (62%), MAI – KZT 3.21 million (13%), distribution of milk or equivalent food products – KZT 2.13 million (8%).

It should be noted that the nature of OSH expenditures is determined by the classification of working conditions as hazardous and/or harmful. Accordingly, the total volume of OSH expenditures—including the dominant categories listed above—is calculated based on the number of employees engaged in such conditions, as defined by the approved occupational risk lists.

Thus, expenditures for the provision of additional paid leave are calculated based on the number of actual working days performed under hazardous and/or harmful conditions, taking into account the employee's occupation in accordance with the approved list, with a minimum of six calendar days. MPPC expenditures are calculated as 5% of the wage fund for employees engaged in heavy work or work under hazardous conditions.

MAI expenditures are calculated as 1.55% of the total wage fund of the enterprise, based on the classification of its economic activity under Professional Risk Class 14, which determines the amount of the insurance premium under the insurance contract.

A comparative analysis of OSH expenditure dynamics reveals a baseline growth rate of 9.7%. The increase is primarily driven by rising costs in the following areas: periodic medical examinations – up by 229.1%, pre-shift medical screenings – up by 120.0%, and advanced training for managers and OSH-responsible personnel – up by 73.7%. These increases are attributed to the growth in the number of employees at the enterprise.

The OSH expenditure analysis for **Enterprise D**, which employs a total of 1,447 workers – 1,131 of whom (78%) are engaged in hazardous working conditions – shows that, during the analyzed period, funds were allocated to 10 out of 17 OSH expenditure categories. Over the three years, the dominant OSH expenditure items were: increased wage payments for employees engaged in heavy work or work under hazardous and/or dangerous conditions – KZT 106.07 million (36%), provision of additional paid leave for such employees – KZT 59.6 million (20%), MPPC – KZT 37.4 million (13%), MAI – KZT 38.01 million (13%). It should be noted that the structure of OSH expenditures reflects the classification of working conditions as hazardous and/or harmful.

MAI expenditures are calculated as 1.13% of the total wage fund of all employees, based on the classification of the enterprise's economic activity under Professional Risk Class 15, which determines the insurance premium under the contract.

A comparative analysis of OSH expenditure dynamics shows a baseline growth rate of 46.1%. The increase is primarily due to rising costs in the following areas: provision of additional paid leave for employees working under hazardous and/or dangerous conditions, up by 31.2%, periodic medical examinations, up by 66.1%, training, instruction, and knowledge testing of employees, up by 28.2%. Overall, this growth is associated with a 25% increase in the number of employees working under hazardous and/or harmful conditions. At the same time, there was a notable decrease in expenditures for: advanced training for managers and OSH-responsible personnel – down by 62.6%, which is explained by the training cycle for managers (every three years); sanitary and welfare services – down by 30.3%.

The OSH expenditure analysis for **Enterprise E**, which employs a total of 4,102 workers – 3,484 of whom (84%) work under hazardous conditions – shows that, during the analyzed period, funds were allocated to 15 out of 17 OSH expenditure categories. Over the three years, the dominant OSH expenditure items were: increased wage payments for employees – KZT 3,100 million (46%), provision of PPE – KZT 1,063.8 million (16%), distribution of therapeutic and preventive nutrition (TPN) and vitamin supplements – KZT 544 million (8%), reduced working hours – KZT 517 million (8%). The structure of OSH expenditures reflects the classification of working conditions as hazardous and/or harmful. Accordingly, the total volume of OSH expenditures—including the dominant categories listed above—is determined based on the number of employees engaged in hazardous and/or harmful conditions, by the approved lists. MAI expenditures are calculated as 1.55% of the total

wage fund of all employees, based on the classification of the enterprise's economic activity under Professional Risk Class 14, which determines the insurance premium under the insurance contract.

A comparative analysis of OSH expenditure dynamics shows a baseline growth rate of 37.6%. The increase is primarily driven by rising expenditures in the following areas: sanitary and welfare services – up by 149.8%, equipping OSH and safety training rooms – up by 303.9%, distribution of therapeutic and preventive nutrition and vitamin supplements – up by 137.1%,
provision of PPE, cleansing and neutralizing agents – up by 79.9%. These increases are associated with a 17.9% growth in the number of employees.

The OSH expenditure analysis for **Enterprise F**, a small enterprise with a total workforce of 58 employees, indicates that no hazardous working conditions were identified based on the results of workplace certification. During the analyzed period, funds were allocated to 8 out of 17 OSH expenditure categories. The dominant OSH expenditure items were: provision of PPE – KZT 1.73 million (42%), MAI – KZT 0.86 million (21%), sanitary and welfare services – KZT 0.82 million (20%). It is worth noting that, during the inspection, non-compliance with the requirement to conduct pre-shift medical screenings was identified, which was attributed to insufficient OSH funding.

The OSH expenditure analysis for **Enterprise G**, which employs a total of 247 workers—207 of whom (83.8%) are engaged in hazardous working conditions—shows that, during the analyzed period, funds were allocated to 9 out of 17 OSH expenditure categories. Over the three years, the dominant OSH expenditure items were: increased wage payments for employees – KZT 41 million (27%), provision of PPE, cleansing and neutralizing agents – KZT 31.86 million (20%), distribution of milk or equivalent food products – KZT 27.8 million (18%).

The structure of OSH expenditures reflects the classification of working conditions as hazardous and/or harmful. A comparative analysis of OSH expenditure dynamics reveals a baseline growth rate of 9.3%.

The analysis of OSH expenditures across the surveyed enterprises over three years demonstrates the stability of cost structures in terms of their dominant categories. This consistency confirms the alignment between the nature of the expenditures and each enterprise's OSH management strategy. The dominant OSH cost categories across the enterprises were: increased wage payments – 22%, additional paid leave – 21%, provision of PPE – 20%, MPPC – 12%.

These expenditures are justified by the significant proportion of employees engaged in hazardous working conditions.

Specialized software tools can be used for analyzing OSH expenditures. Below is a summary of the functionality of OSHA's «Safety Pays» program [20], developed by the Occupational Safety and Health Administration (U.S. Department of Labor). This tool provides information on the frequency and cost of occupational injuries and illnesses.

Based on data collected from thousands of workplaces, «Safety Pays» offers a unique opportunity to benchmark a company's workplace safety indicators against those of similar enterprises. The tool also helps evaluate compliance with OSH obligations.

Employers can utilize the «Safety Pays» tool in various ways, depending on the extent of available data on occupational injuries and diseases within their organization. The tool is particularly effective when used in conjunction with OSHA Form 300A, which employers are required to complete. In the absence of such data, the tool can still be used to estimate the potential financial burden associated with workplace injuries and illnesses.

The functionality of the «Safety Pays» program, in the absence of internal occupational injury and illness statistics, allows users to:

- estimate the probability of injury based on the industry and workforce size;
- assess the potential costs associated with a typical injury within the selected industry;
- identify current expenditures and evaluate their impact on business operations.

When OAD data are available, this tool can be used to:

- compare the company's insurance rate with those of similarly sized employers within the same industry;
- calculate the potential annual burden associated with these cases.

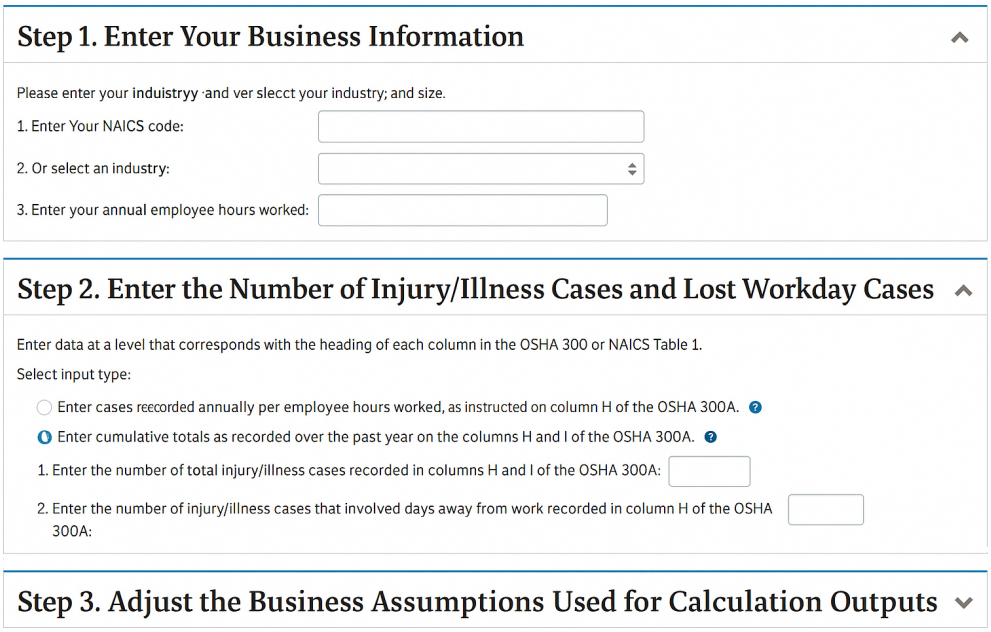

Figure 2 – Functionality of the «Safety Pays» Program

The «Safety Pays» tool uses the following data sources in its calculations:

1. Occupational injury and illness rates by industry.

OSHA utilizes industry-specific injury quartile tables from the Bureau of Labor Statistics (BLS) to determine or calculate the Total Case Rate (TCR) and the DART rate (Days Away, Restricted, or Transferred) across industries and establishment size categories.

The «Safety Pays! tool applies two benchmark estimates:

- the average rate for the entire industry and establishment size category («all competitors»),
- the average rate among firms below the 25th percentile («top-performing competitors»).

2. Types of Injuries and Illnesses by Industry. The Bureau of Labor Statistics (BLS) provides data on injuries and illnesses by nature of injury or illness (e.g., fracture) for serious nonfatal cases (i.e., cases where the worker was away from work for at least one day beyond the day of injury or illness onset). OSHA uses this dataset to calculate the proportion of serious cases by injury or illness type for specific industry sectors.

3. Costs of Wage-Loss Compensation Claims by Type of Injury. The cost estimates for wage-loss compensation claims used in the «Safety Pays» tool are provided by the National Council on Compensation Insurance, Inc. (NCCI). These data reflect the average cost of wage-loss claims based on statistical reports from member jurisdictions submitted to NCCI.

It is important to note that in the United States, NCCI manages the nation's largest database of workers' compensation insurance information. NCCI analyzes industry trends, provides rate recommendations for workers' compensation insurance, evaluates the cost of proposed legislation, and offers various services and tools to support the effective operation of the workers' compensation system.

Significantly, the estimates of direct (i.e., workers' compensation) and indirect costs are provided in dollar terms across major industry sectors. Workers' compensation claims may include medical-only claims (covering medical expenses only) and lost-time claims (covering wage-loss payments due to time away from work).

3. For medical-only claims, the tool uses the average cost of an emergency department visit, based on data from the Centers for Disease Control and Prevention (CDC) Web-based Injury Statistics Query and Reporting System (WISQARS). This cost estimate applies to all nonfatal injuries that are treated and released (i.e., not resulting in hospitalization) for individuals aged 20 to 64. In addition

to lost-time claims, a company's workers' compensation costs may also include medical-only claims. For such claims, the tool utilizes the average cost of an emergency department visit, derived from the CDC's Web-based Injury Statistics Query and Reporting System (WISQARS). This cost estimate applies to all nonfatal injuries that are treated and released (i.e., not resulting in hospitalization) for individuals aged 20 to 64.

4. Indirect Costs. The estimates of indirect costs presented in this tool are drawn from the publication Improving Construction Safety Performance by the Business Roundtable and are based on a study conducted by the Department of Civil Engineering at Stanford University.

Table 2 – Indirect Costs

Direct Costs	Indirect Cost Ratio
$0–2,999	4.5
$3,000–4,999	1.6
$5,000–9,999	1.2
$10,000 or more	1.1

The magnitude of indirect costs is inversely proportional to the severity of the injury: the less severe the injury, the higher the ratio of indirect to direct costs. Although indirect costs may constitute a significant portion of the actual cost of an incident, they are typically uninsured and therefore not reimbursable.

The indirect cost multipliers used in the Safety Pays tool represent average estimates (across all injury types, based on data from the construction industry) and are derived from a limited study on this subject. The actual indirect cost multiplier may vary depending on the employer's specific circumstances.

Below are cost estimates by type of OAD, as well as by industry sector.

Table 3 – Cost estimates by type of OAD

Type of Injury	Cost*	Type of Injury	Cost*
Asphyxia	$221 596	All other cumulative trauma disorders	$46 179
Circulatory disorders	$208 441	Respiratory disorders (gases, fumes, chemicals, etc.)	$44 771
Cancer	$178 367	Inflammation	$42 707
Laceration	$177 312	Heat stroke	$41 109
Electric shock	$172 716	AIDS	$39 973
Silicosis	$163 583	Vision-related disorders	$39 576
Multiple injuries, including both physical and psychological	$133 919	Mental stress	$38 372
Enucleation (e.g., removal of a tumor, eye, etc.)	$112 238	Dust disease, NOC (not otherwise classified pneumoconioses)	$38 109
Amputation	$104 800	Overexertion	$34 957
Multiple physical injuries only	$85 301	Infection	$34 278
Loss of vision	$82 838	Carpal tunnel syndrome	$33 764
Dislocation	$82 080	Sprain	$33 281
Fracture	$74 947	Fainting	$33 229
Angina	$73 775	Metal poisoning	$32 056
Asbestosis	$73 496	Puncture	$30 352
Crushing injury	$73 143	Contusion	$30 162
Myocardial infarction (heart attack)	$72 598	Chemical poisoning (excluding metals)	$29 782
Mental disorder	$64 863	Non-physical injury	$28 233
Radiation	$63 910	Frostbite	$25 193
Pneumoconiosis	$63 883	Hearing loss or impairment (injury-related only)	$24 861
Fracture (repeated for clarity if needed)	$59 882	Foreign body	$23 971
Concussion	$59 571	Avulsion	$23 876
All other specific injuries	$58 899	Hernia	$22 859
All other occupational illnesses	$51 583	Hearing loss	$20 553
Burn	$51 516	Infectious disease	$11 777
Poisoning – general (not cumulative trauma)	$50 703	Dermatitis	$10 568

*Average cost of a lost-time claim

Indirect cost estimates are calculated by applying a multiplier to the direct cost estimates [2].

[2]These multipliers are based on a 1982 study conducted by researchers at Stanford University.

Table 4 – Cost estimates related to OAD by industry sector

Code	Sector Name	Direct Lost-Time Costs	Indirect Lost-Time Costs	Direct Medical-Only Costs	Indirect Medical-Only Costs
00X	All Industries	$43 920	$48 311	$3 130	$14 083
ADW	Administrative, Support, and Waste Management Services	$42 922	$47 214	$3 130	$14 083
AER	Arts, Entertainment, and Recreation	$43 666	$48 033	$3 130	$14 083
AFH	Agriculture, Forestry, Fishing, and Hunting	$45 514	$50 065	$3 130	$14 083
AFS	Accommodation and Food Services	$42 136	$46 349	$3 130	$14 083
EDS	Educational Services	$44 714	$49 185	$3 130	$14 083
FIN	Finance and Insurance	$49 626	$54 589	$3 130	$14 083
HAS	Health Care and Social Assistance	$42 749	$47 024	$3 130	$14 083
MCE	Management of Companies and Enterprises	$44 465	$48 911	$3 130	$14 083
MIN	Mining, Quarrying, and Oil and Gas Extraction	$43 889	$48 278	$3 130	$14 083
PST	Professional, Scientific, and Technical Services	$41 345	$45 480	$3 130	$14 083
RET	Retail Trade	$42 015	$46 216	$3 130	$14 083
RRL	Real Estate and Rental and Leasing	$49 046	$53 950	$3 130	$14 083
TRW	Transportation and Warehousing	$42 008	$46 208	$3 130	$14 083
UTL	Utilities	$46 407	$51 048	$3 130	$14 083
WHT	Wholesale Trade	$43 099	$47 409	$3 130	$14 083
CON	Construction	$45 671	$50 238	$3 130	$14 083
MFG	Manufacturing	$45 068	$49 575	$3 130	$14 083
INF	Information	$39 962	$43 958	$3 130	$14 083
OTS	Other Services (except Public Administration)	$43 181	$47 499	$3 130	$14 083
SP1	Public Administration	$42 961	$47 257	$3 130	$14 083

Thus, the OSHA «Safety Pays» tool is an online calculator that utilizes injury and illness cost data collected by the National Council on Compensation Insurance (NCCI). It processes this data on your behalf to calculate the direct and indirect costs of an injury, as well as the additional revenue your business would need to generate in order to offset those costs.

According to OSHA, the «Safety Pays» program has been used in the United States to estimate the economic losses resulting from workplace fatalities, injuries, and illnesses. For example, in 2021, it was reported that employers paid over $1 billion per week in direct workers' compensation costs for disabling, nonfatal workplace injuries that occurred in 2018. The National Safety Council estimated that work-related deaths and injuries cost the nation, employers, and individuals approximately $171 billion in 2019. Employers who implement effective OSH programs can expect to significantly reduce the number of workplace injuries and illnesses, as well as the associated costs, including workers' compensation claims, medical expenses, and productivity losses. Moreover, employers often find that the process improvements and operational changes introduced to enhance OSH performance lead to substantial gains in productivity and profitability for their organizations.

4. CONCLUSION

The findings presented in this article are interim results and will be further developed with the aim of designing a software tool that will enable the calculation of required funding (budgeting) for OSH measures, as well as the accounting and analysis of related expenditures.

Numerous studies have demonstrated that investments in OSH yield tangible benefits from a microeconomic perspective, with a return on prevention ratio of 2.2. In practical terms, this means that for every KZT 1 invested annually by companies in prevention, there is a potential economic return of KZT 2.20. In this regard, the authors of the article aim to produce scientifically grounded evidence, through the continuation of this research, in support of corporate investment in preventive OSH measures.

Against the backdrop of the widespread adoption of digital technologies and automation systems, there is a growing need for specialized software to monitor and analyze occupational risks, as well as the development of online platforms for training and informing workers about safety measures. It is also important to establish economic incentives for employers. This may include the design and implementation of financial benefits and tax preferences for enterprises that demonstrate a high level of workplace safety.

ACKNOWLEDGMENTS

The scientific research was based on primary data collected from enterprises, which enabled the formation of an information-analytical database for the accounting and analysis of OSH expenditures.

The authors express their sincere appreciation for the support provided in gathering the necessary information while ensuring full confidentiality.

The scientific results were obtained within the framework of the research project titled «Improvement of the System of Accounting and Analysis of OSH Expenditures in High-Risk Industries» (Project No. AP19680581), funded under the grant program for scientific and/or scientific-technical projects of the Ministry of Science and Higher Education of the Republic of Kazakhstan.

Authors' Contributions:

Sh. Abikenova developed the methodological framework of the study, identified the scientific novelty, and substantiated the use of mathematical methods for analyzing occupational safety expenditures. She also provided the overall scientific editing of the article.

Sh. Aitimova conducted the empirical part of the research, including the collection, systematization, and comparative analysis of occupational safety expenditure data across industries with high rates of occupational injuries. She also contributed to the development of the analytical data bank.

A. Kocherbayeva participated in the interpretation of the relationship between economic losses and investments in preventive measures and contributed to the formulation of conclusions and practical recommendations.

References

1 О ратификации Конвенции об основах, содействующих безопасности и гигиене труда (Конвенция 187): Закон Республики Казахстан от 20 октября 2014 года № 243-V ЗРК.

2 Об утверждении Концепции безопасного труда Республики Казахстан на 2024-2030 годы: Постановление Правительства Республики Казахстан от 28 декабря 2023 года № 1182.

3 Охрана труда в экономическом контексте: научно-информационное издание. Под ред. к.ф.м.н., ассоц. профессора Абикеновой Ш.К. - Астана: РГП на ПХВ «РНИИОТ МТСЗН РК», 2024. - 171 с., ТОО «Полиграф-мир».

4 Трудовой кодекс от 23 ноября 2015 года № 414-V ЗРК. [Электронный ресурс]. -URL: https://adilet.zan.kz/rus/docs/K1500000414

5 Свидетельство авторского права на Методику анализа затрат на охрану труда №48742 от 31.07.2024.

6 Об утверждении Правил ведения бухгалтерского учета: Приказ Министра финансов Республики Казахстан от 31 марта 2015 года № 241.

7 О налогах и других обязательных платежах в бюджет (Налоговый кодекс): Кодекс Республики Казахстан от 25 декабря 2017 года № 120-VI ЗРК.

8 Rikhardsson Pall. Accounting for Health and Safety Costs. Review and Comparison of Selected Methods. University of Aarhus, Aarhus School of Business, Department of Business Studies, Management Accounting Research Group Working Papers. 2025. DOI: 10.1007/978-1-4020-4974-3 6.

9 Mustard C., Yanar B. Estimating the Financial Benefits of Employers' Occupational Health and Safety Expenditures // Safety Science. 2023. Vol. 159(2). 106008.

10 Wirahadikusumah, Reini & Adhiwira, Felix. The cost of implementing OHSMS regulation on high-rise building projects // MATEC Web of Conferences. -2007. 270. 05007. DOI: 10.1051/matecconf/201927005007.

11 Павлова В.Л. Бюджетирование безопасности трудовых процессов на железнодорожном транспорте: дис. канд. эконом. наук. – Новосибирская государственная академия водного транспорта, 2006.

12 Estimating the Economic Costs of Occupational Injuries and Illnesses in Developing Countries: Essential Information for Decision-Makers, First published 2012. -URL: https://www.ilo.org/publications/estimating-economic-costs-occupational-injuries-and-illnesses-developing

13 Heinrich H.W. Industrial Accident Prevention. A Scientific Approach. McGraw-Hill, New York, 4th edition. -1959.

14 Grimaldi J.V., Simonds R.H. Safety Management. New York: Richard D. Irwin -1984.

15 Larsson T.J., Betts N.J. The Variation of Occupational injury Cost in Australia; Estimates Based on a Small Empirical Study. Safety Science. - 1996. -Vol. 24. -No. 2. -143-155

16 Monnery N. The costs of accidents and work-related ill health to a cheque-clearing department of a financial services organisation // Safety Science. - 1998. -Vol. 31. -P. 59-69.

17 Andreoni D. The Cost of Occupational Accidents and Diseases. International Labour Office, Geneva. -1986.

18 Hrymak V., Perezgonzalez J. The costs and effects of workplace accidents: 20 case studies from Ireland. -2007.

19 Dale B.G., Plunkett J.J. Quality Costing. -London: Chapman & Hall/ 2007.

20 Safety Pays Tool Introduction. -URL: https://www.osha.gov/safetypays/tool (Accessed: 17.01.2025)

Еңбек қорғау шығындарын есепке алу және талдау бойынша ақпараттық-талдамалық деректер базасын қалыптастырудың негізі мен ғылыми дәйегі ретіндегі математикалық әдістер

Ш. Әбікенова [1], Ш. Айтимова [2], А. Кочербаева [3]
[1] «ҒАЛЫМ» Ғылыми-зерттеу компаниясы» ЖШС, Ғ.Мүсірепов көш.,15а, Астана, Қазақстан Республикасы
[2] «Мемлекеттік әлеуметтік сақтандыру қоры» АҚ, Достық көш., 18, Астана, Қазақстан Республикасы
[3] Б.Н. Ельцин атындағы Қырғыз-Ресей Славян университеті, Киевская көш., 44, Бішкек, Қырғызстан

Аннотация. Қазіргі жағдайда өндірістік жарақаттар мен кәсіби аурулардан туындайтын экономикалық шығындарды түрлі экономикалық салалар аясында зерттеуге қызығушылық артып отыр. Бұл бағыттағы негізгі әдістердің бірі – құрылымдық, серпінді және салыстырмалы есептеулерде қолданылатын математикалық әдістер. Экономикалық шығындар мен еңбек қорғауға (превентивтік шығындарға) салынатын инвестициялар арасындағы кері пропорционалды өзара байланыс – қызметкер денсаулығын сақтау, еңбек өнімділігін арттыру және жалпы экономикалық тиімділікті қамтамасыз етуде маңызды фактор болып табылады. Осыған орай, еңбек жағдайларын жақсартуға бағытталған шығындарды салыстырмалы түрде бағалау – еңбек қорғау жүйесін жетілдірудің маңызды құралдарының бірі ретінде қарастырылады.

Мақалада еңбек қорғау саласындағы шығындарды есепке алу және талдау мақсатында ақпараттық-талдамалық деректер базасын қалыптастырудың ғылыми негіздері қарастырылады. Бұл үшін математикалық әдістердің ішінде салыстыру, топтастыру және жүйелеу тәсілдері қолданылып, бастапқы деректер еңбек қорғауға жұмсалатын негізгі шығын баптары бойынша салыстырылған. Талдауға жататын шығын түрлері Қазақстан Республикасының еңбек заңнамасына сәйкес айқындалған. Зерттеудің мақсаты – еңбек қорғау шығындарын есепке алу мен талдауға арналған бағдарламалық өнім әзірлеу барысында математикалық әдістерді қолдану.

Түйін сөздер: математикалық әдістер, ақпараттық-талдамалық деректер базасы, еңбек қорғау шығындарын есепке алу және талдау, базалық көрсеткіштер, бағдарламалық өнім.

Математические методы как основа и научное обоснование к формированию информационно-аналитического банка данных по учету и анализу затрат на охрану труда

Ш. Абикенова [1], Ш. Айтимова [2], А. Кочербаева [3]
[1] ТОО «Научно-исследовательская компания «ҒАЛЫМ», ул. Г.Мусрепова 15а, Астана, Республика Казахстан
[2] АО «Государственный фонд социального страхования», ул. Достык, 18, Астана, Республика Казахстан
[3] Кыргызско-Российский Славянский университет имени Б.Н. Ельцина, ул. Киевская, 44, Бишкек, Кыргызская Республика

Аннотация. В современных условиях возрастает интерес к исследованию экономических потерь от производственного травматизма и профзаболеваний в различных отраслях экономики. Основными методами исследования являются математические методы, которые используются при расчетах, анализе структуры, динамики и т.д. Влияние экономических потерь и их обратно-пропорциональная связь с инвестициями в охрану труда (превентивными расходами) является важным в сохранении здоровья работника, повышения производительности труда и экономики в целом, что обосновывает необходимость в совершенствовании охраны труда, посредством разнообразных инструментов, одним из которых является сравнительный аспект затрат на охрану труда предприятия, в частности, направленных на улучшение условий труда в анализируемом периоде.

В статье дается обоснование информационно-аналитического банка данных по учету и анализу затрат на охрану труда на основе применения математических методов, среди которых отметим методы сравнения, группировки и систематизации посредством сопоставления первичных данных по основным статьям расходов на охрану труда. Виды анализируемых расходов определены согласно нормам трудового законодательства Республики Казахстан. Цель исследования состоит в применении математических методов при разработке программного продукта по учету и анализу затрат на охрану труда.
Ключевые слова: математические методы, информационно-аналитический банк данных, учет и анализ затрат на охрану труда, базовые значения, программный продукт.

References

1 O ratifikatsii Konventsii ob osnovakh, sodeystvuyushchikh bezopasnosti i gigiene truda (Konventsiya 187): Zakon Respubliki Kazakhstan ot 20 oktyabrya 2014 goda № 243-V ZRK [On the Ratification of the Promotional Framework for Occupational Safety and Health Convention (Convention No. 187): Law of the Republic of Kazakhstan No. 243-V ZRK dated October 20, 2014,]. [Electronic resource]. Available at: https://adilet.zan.kz/rus/docs/Z1400000243 [in Russian].

2 Ob utverzhdenii Kontseptsii bezopasnogo truda Respubliki Kazakhstan na 2024–2030 gody: Postanovlenie Pravitel'stva Respubliki Kazakhstan ot 28 dekabrya 2023 goda № 1182 [On the Approval of the Concept for Safe Labor in the Republic of Kazakhstan for 2024–2030: Resolution of the Government of the Republic of Kazakhstan No. 1182 dated December 28, 2023]. [Electronic resource]. Available at: https://adilet.zan.kz/rus/docs/P2300001182 (in Russian).

3 Nauchno-informatsionnoe izdanie «Okhrana truda v ekonomicheskom kontekste», pod red. k.f.m.n., assots. professora Abikenovoy Sh.K. – Astana: RGP na PHV "RNIIOT MTSZN RK", 2024. – 171 s., TOO "Poligraf-mir", [Scientific and informational publication Occupational Safety and Health in the Economic Context, edited by Sh.K. Abikenova, PhD in Physics and Mathematics, Associate Professor. – Astana: RSE on REM «Republican Research Institute for Occupational Safety of the MLSPP RK», 2024. – 171 p. – TOO «Poligraf-Mir»]

4 Trudovoy kodeks ot 23 noyabrya 2015 goda № 414-V ZRK [Labour Code of the Republic of Kazakhstan, No. 414-V ZRK, dated November 23, 2015] [Electronic resource]. Available at: https://adilet.zan.kz/rus/docs/K1500000414 [in Russian].

5 Svidetel'stvo avtorskoho prava na Metodiku analiza zatrat na okhranu truda №48742 ot 31.07.2024 [Certificate of Copyright for the Methodology for Analyzing Occupational Safety and Health Expenditures, No. 48742, dated July 31, 2024] [in Russian].

6 Ob utverzhdenii Pravil vedeniya bukhgalterskogo ucheta: Prikaz Ministra finansov Respubliki Kazakhstan ot 31 marta 2015 goda № 241 [On the Approval of the Rules for Maintaining Accounting Records. Order of the Minister of Finance of the Republic of Kazakhstan dated March 31, 2015, No. 241.] [Electronic resource]. Available at: https://adilet.zan.kz/rus/docs/V1500010954 [in Russian].

7 O nalogakh i drugikh obyazatel'nykh platezhakh v byudzhet (Nalogovyy kodeks), Kodeks Respubliki Kazakhstan ot 25 dekabrya 2017 goda № 120-VI ZRK [On taxes and other obligatory payments to the budget (Tax Code). Code of the Republic of Kazakhstan of December 25, 2017 № 120-VI] [Electronic resource]. Available at: https://adilet.zan.kz/eng/docs/K1700000120

8 Rikhardsson, Pall. (2005). Accounting for Health and Safety Costs. Review and Comparison of Selected Methods. University of Aarhus, Aarhus School of Business, Department of Business Studies, Management Accounting Research Group Working Papers. DOI: $10.1007/978-1-4020-4974-3_6$.

9 Mustard C., Yanar B. Estimating the Financial Benefits of Employers' Occupational Health and Safety Expenditures, Safety Science. 2023. Vol. 159(2). 106008.

10 Wirahadikusumah, Reini & Adhiwira, Felix. The cost of implementing OHSMS regulation on high-rise building projects. MATEC Web of Conferences. 2019. 270. 05007. DOI: 10.1051/matecconf/201927005007.

11 Pavlova V.L. Byudzhetirovanie bezopasnosti trudovykh protsessov na zheleznodorozhnom transporte: dis. – Novosibirskaya gosudarstvennaya akademiya vodnogo transporta [Budgeting for the safety of labor processes in railway transport (Doctoral dissertation). Novosibirsk State Academy of Water Transport]. 2006. [in Russian].

12 Estimating the Economic Costs of Occupational Injuries and Illnesses in Developing Countries: Essential Information for Decision-Makers, First published 2012. Available at: https://www.ilo.org/publications/estimating-economic-costs-occupational-injuries-and-illnesses-developing

13 Heinrich H.W. Industrial Accident Prevention. A Scientific Approach. McGraw-Hill, New York, 4th edition. 1959.

14 Grimaldi J.V., Simonds R.H. Safety Management. New York: Richard D. Irwin. 1984.

15 Larsson T.J., Betts N.J. The Variation of Occupational injury Cost in Australia; Estimates Based on a Small Empirical Study, Safety Science. 1996. Vol. 24. No. 2. P. 143-155

16 Monnery N. The costs of accidents and work-related ill health to a cheque-clearing department of a financial services organisation, Safety Science. 1998. Vol. 31. P. 59-69.
17 Andreoni D. The Cost of Occupational Accidents and Diseases. International Labour Office, Geneva. 1986.
18 Hrymak V., & Perezgonzalez J. The costs and effects of workplace accidents: 20 case studies from Ireland. 2007.
19 Dale B.G., Plunkett J.J. Quality Costing. London: Chapman & Hall. 1997.
20 Safety Pays Tool Introduction. Available at: https://www.osha.gov/safetypays/tool (Accessed: 17.01.2025)

Information about authors:

Sholpan Abikenova – cand. of phys.-math. sci., Associate Professor, «Scientific Research Company GALYM» LLP, G.Musrepov str., 15a, Astana, Republic of Kazakhstan.

Shynar Aitimova – Chief Manager of the Project Office, JSC «State Social Insurance Fund», Dostyk str., 18, Astana, Republic of Kazakhstan.

Ainura Kocherbaeva – Doctor of Economics, Professor, Dean of the Faculty of Economics Kyrgyz-Russian Slavic University named after B. N. Yeltsin, Kievskaya str., 44, Bishkek, Kyrgyzstan.

Шолпан Кәкімжанқызы Әбікенова – ф,-м.ғ.к., доцент, «ҒАЛЫМ» Ғылыми-зерттеу компаниясы» ЖШС, Ғ.Мүсрепов көш., 15а, Астана, Қазақстан Республикасы.

Шынар Тұрсынханқызы Айтимова – бас менеджер, «Мемлекеттік әлеуемттік сақтандыру қоры» АҚ жобалық кеңсесі, Достық көш., 18, Астана, Қазақстан Республикасы.

Айнура Анатольевна Кочербаева – э.ғ.к., профессор, экономика факультетінің деканы, Б.Н. Ельцин атындағы Қырғыз-Ресей Славян университеті, Киевская көш., 44, Бішкек, Қырғызстан.

Received: 25.01.2025. Revised: 23.02.2025.
Approved: 19.03.2025. Available online: 31.03.2025.

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (https://creativecommons.org/licenses/by-nc/4.0/).

Бас редактор: Н. Темірғалиев

Жауапты редактор: Ғ.Е. Тауғынбаева

Л.Н. Гумилев атындағы Еуразия ұлттық университетінің
хабаршысы. Математика, компьютерлік ғылымдар, механика сериясы.
- 2025. 1(150). - Астана: ЕҰУ. 42-б. Басуға қол қойылды: 30.03.2025.
Ашық қолданыстағы 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