sonli modeli qurilgan. Murakkab osesimmetrik jismning ikki o‘lchovli kesimini diskret modelini qurish va masalani yechish algoritmi ishlab chiqildi. 1-masalani yechish orqali olingan sonli natijalar Elcut dasturidan olingan natijalar bilan taqqoslandi va yaratilgan algoritm to‘g‘ri ekanligi isbotlandi. Bir jinsli bo‘lmagan osesimmetrik jism geometrik xususiyatlarning harorat maydoniga ta’siri o‘rganildi va issiqlik oqimi strukturaning tanasida to‘rtburchaklar yoki aylana kesmalarning silindrsimon bo‘shlig‘i mavjudligida tahlil qilindi.

Adabiyotlar

1. Зенкевич О. Метод конечных элементов в технике. – М.: «Мир», 1975. 452-b.
2. Сегерлинд Л. Применение метода конечных элементов. – М.: «Мир», 1979. 392-b.
3. Иванников Л. М. Решение двумерной задачи теплопроводности методом конечных элементов в Mathcad. Хабаровск: Тихоокеан. гос. ун-та, 2011. -28 b.
4. Икрамов А.М., Жуманиёзов С.П., Сапаев Ш.О. Компьютерное моделирование двумерных стационарных задач теплопроводности с учетом точечных источников тепла МКЭ. //Проблемы вычислительной и прикладной математики, 2021, 3 (33). –С. 44-53.

REZYUME. Maqolada ChEU ga asoslanib, osesimmetrik jismlarda issiqlik tarqalishining nostatsionar masalalarini yechishning sonli modeli ishlab chiqilgan, masalani yechish algoritmi va dasturiy ta’minoti yaratilgan.

РЕЗЮМЕ. В статье на основе ЧЭУ разработана численная модель решения нестационарной задачи распределения тепла в осесимметричных телах, созданы алгоритм и программное обеспечение для решения задачи.

SUMMARY. In the article, based on FEM, a numerical model for solving non-stationary problems of heat dissipation in axisymmetric bodies is developed, and an algorithm and software for solving the problem are created.

РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ ДЛЯ КАРАКАЛПАКСКОГО ЯЗЫКА С ИСПОЛЬЗОВАНИЕМ Wav2Vec

Н.У.Утеулиев – *доктор физико-математических наук, профессор*
Ж.К.Кудайбергенов – *докторант*
Т.К.Кудайбергенов – *магистрант*
Нукусский филиал Ташкентского университета информационных технологии

Таянч сўзлар: ASR, Wav2Vec**,** таниб олиш, модел, частота, метрика.
Ключевые слова: ASR, Wav2Vec, распознавание, модель, частота, метрика.
Key words: ASR, Wav2Vec, recognition, model, frequency, metrics.

Введение. В настоящее время растет число исследований, посвященных автоматической обработке малоресурсных языков. Следует отметить, что современные технологии обработки естественного языка активно развиваются и находят применение в различных областях, включая автоматическое разпознавание речи. Но отсутствие развитых речевых технологий для малоресурсных языков остаётся актуальной проблемой сегодняшнего дня. Системы автоматического распознавания речи (ASR) достигли значительного прогресса в последние годы, преимущественно для широко распространенных языков. Однако многие языки, такие как каракалпакский, по-прежнему остаются недостаточно представленными в исследованиях ASR. Предлагаемая статья направлена на устранение этой проблемы путем разработки структуры сбора данных, которая захватывает аудио экземпляры и их соответствующие текстовые представления. Используя модель Wav2Vec, мы стремимся создать эффективную систему ASR, которая переводит устный каракалпакский язык в письменный текст.

Каракалпакский язык [3] – язык каракалпакского народа, это один из тюркских языков, который используется в основном в Узбекистане. Несмотря на его культурное значение, язык не имеет достаточных цифровых ресурсов, что делает его сложной целью для разработки ASR.

Wav2Vec – это фреймворк само обучаемый вид обучения, разработанный Facebook AI Research, который произвел революцию в области распознавания речи. Модель работает, обучаясь представлениям сырых звуковых волн, что позволяет ей захватывать нюансы речи без необходимости в больших размеченных наборах данных. Wav2Vec использует двухступенчатый процесс обучения:

Предварительное обучение: На этом этапе модель обучается предсказывать замаскированные сегменты аудио из окружающего контекста, используя большое количество неаннотированных аудио данных. Этот шаг помогает модели понять фонетические и лингвистические особенности, присущие речи.

Тонкая настройка: После предварительного обучения модель проходит тонкую настройку на меньшем размеченном наборе данных, где она учится сопоставлять аудио представления с текстом. Этот подход трансферного обучения значительно снижает количество необходимых размеченных данных, что делает его особенно эффективным для языков с ограниченными ресурсами, таких как каракалпакский. Wav2Vec показал передовые результаты по различным стандартам, превосходя традиционные методы ASR. Его архитектура включает свёрточные слои, ко-торые извлекают признаки из аудиовхода, за которыми следуют слои Transformer, которые захватывают дальние зависимости в данных. Это сочетание позволяет Wav2Vec обрабатывать разнообразные речевые паттерны, что делает его подходящим для языков с богатым фонетическим разнообразием.

Сбор данных. Первым шагом было создание Java-приложения для захвата аудио потоков с микрофона. Приложение читает предложения построчно из заранее определенного текстового файла и записывает аудио в формате WAV.

Реализация захвата аудио. Java-код использует пакет javax.sound.sampled для захвата аудио. Каждое

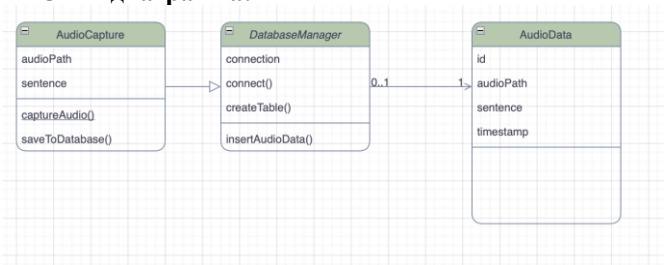
записанное предложение сохраняется как отдельный WAV-файл, обеспечивая высокое качество аудио для последующей обработки.

```
// Фрагмент Java-кода для захвата аудио
import javax.sound.sampled.*;
public class AudioCapture {
  public void captureAudio(String sentence) {
    // Логика захвата аудио
  }
}
```

Хранение в базе данных. Для обеспечения эффективного управления данными каждый аудиофайл связывается с соответствующим предложением в базе данных SQLite. Схема базы данных включает поля для пути к аудиофайлу, текста предложения и временных меток.

```
CREATE TABLE audio_data (
id INTEGER PRIMARY KEY,
audio_path TEXT NOT NULL,
sentence TEXT NOT NULL,
timestamp DATETIME DEFAULT CURRENT_TIMESTAMP);
```

UML-диаграмма:



Пред обработка: Перед подачей аудио данных в модель Wav2Vec мы предварительно обрабатываем аудиофайлы, чтобы обеспечить единообразие в частотах дискретизации и глубине битов. Этот шаг имеет решающее значение для производительности модели. Этап предварительной обработки включает:

Нормализация: Регулировка уровней громкости аудиофайлов, чтобы предотвратить искажения.

- Обрезка тишины: Удаление ведущих и заключительных пауз для снижения шума в наборе данных.

- Переход на другую частоту дискретизации: Обеспечение того, чтобы все аудиофайлы имели частоту дискретизации 16 кГц, стандарт для модели Wav2Vec.

Аугментация [4] данных: Для повышения устойчивости модели мы реализовали техники аугментации данных. Это включает:

- Изменение скорости: Небольшое изменение скорости воспроизведения аудиофайлов для создания вариаций в наборе данных.

- Инъекция шума: Добавление фонового шума для имитации реальных условий, улучшая способность модели к обобщению.

Реализация. Обучение модели Wav2Vec

Предобработанные аудиоданные используются для тонкой настройки модели Wav2Vec. Процесс обучения включает подачу модели пар аудиофайлов и их соответствующих транскрипций, чтобы научить сопоставлению речи с текстом. Мы использовали трансферное обучение, начав с предобученной модели Wav2Vec и тонко настроив ее на нашем каракалпакском наборе данных.

Метрики оценки. Для строгой оценки производительности системы ASR мы применяем несколько метрик, которые предоставляют информацию о ее точности и надежности:

Коэффициент ошибок слов (WER): WER является стандартной метрикой в ASR, которая количественно оценивает количество ошибок в предсказанной транскрипции по сравнению с эталонной транскрипцией. Он рассчитывается по формуле:

$$WER=(S+D+I)/N$$

где S – количество замен, D – количество удалений, I – количество вставок, а N – общее количество слов в эталонной транскрипции. Например, WER в 5% означает, что 5 из каждых 100 слов были некорректно транскрибированы, что обычно считается приемлемым для многих приложений ASR.

Коэффициент ошибок символов (CER): Подобно WER, CER сосредоточен на точности на уровне символов, что особенно полезно для языков со сложными письменностями или фонетическими вариациями. CER рассчитывается аналогичным образом, но учитывает ошибки на уровне символов. Эта метрика важна при работе с языками, где слова могут иметь множество вариаций в написании и произношении, так как она предоставляет более детальную оценку производительности модели.

Коэффициент реального времени (RTF): RTF измеряет эффективность системы ASR, вычисляя время, затраченное на обработку аудио, относительно длины аудиовхода. Он рассчитывается как:

$$RTF=Processing\ Time/Audio\ Duration.$$

RTF менее 1.0 указывает на то, что система обрабатывает аудио в реальном времени, что критично для практических приложений, таких как живая транскрипция или голосовые команды. Например, RTF 0.5 означает, что система может обрабатывать аудио в два раза быстрее фактического воспроизведения.

Точность, полнота и F1-оценка: Эти метрики предоставляют дополнительную информацию о производительности модели в различении правильных и неправильных транскрипций:

Точность измеряет долю правильно предсказанных слов из всех предсказанных слов, рассчитываемую как:

$$Precision=TP/(TP+FP)$$

где TP – количество истинных положительных (правильно предсказанные слова), а FP – количество ложных положительных (неправильно предсказанные слова). Полнота измеряет долю правильно предсказанных слов из всех фактических слов в эталоне, рассчитываемую как:

$$Recall=TP/(TP+FN)$$

где FN – количество ложных отрицательных (упущенные слова).

F1-оценка – это гармоническое среднее точности и полноты, предоставляющее сбалансированную меру производительности, особенно полезную при неравномерном распределении классов:

$$F1=2*Precision*Recall/(Precision+Recall)$$

Матрица путаницы: Матрица путаницы предоставляет визуальное представление о производительности модели ASR, показывая

количество истинных положительных, ложных положительных, ложных отрицательных и истинных отрицательных для каждого класса (слова). Эта матрица помогает выявить конкретные области, где модель может испытывать трудности, такие как определенные слова или фонемы, которые часто классифицируются неправильно.

Анализируя эти метрики, мы можем получить полное представление о сильных и слабых сторонах системы ASR. Эта оценка не только информирует о последовательных улучшениях модели, но и направляет разработку стратегий для повышения ее устойчивости и точности в реальных приложениях.

Результаты. Предварительные результаты показывают обещающие уровни точности в переводе аудио на каракалпакском языке в текст. Модель достигла WER 12.5% и CER 8.3% на валидационном наборе, демонстрируя потенциал нашего подхода.

Несколько кейс-стадий иллюстрируют производительность системы в различных структурах предложений и фонетических вариациях. Модель проявила устойчивость в распознавании разнообразных речевых паттернов, включая:

Вариации спикеров: Тестирование с записями от нескольких спикеров продемонстрировало адаптивность модели.

Сложные предложения: Модель сохраняла точность даже с более длинными и сложными структурами предложений.

Обратная связь от пользователей:

Первоначальная обратная связь от носителей каракалпакского языка показала, что транскрипции были контекстуально точными, хотя некоторые фонетические нюансы можно улучшить. Эта обратная связь крайне важна для последовательных улучшений модели.

Хотя первоначальные результаты обнадеживают, остаются проблемы, включая диалектные вариации и фоновый шум во время захвата аудио. Будущая работа будет сосредоточена на повышении устойчивости к шуму и расширении набора данных, чтобы включить больше диалектов и контекстов речи.

Дальнейшие исследования будут исследовать интеграцию более продвинутых техник, таких как: Модели "от начала до конца" [5]: Исследование архитектур ASR "от начала до конца", чтобы упростить процесс обучения и потенциально улучшить производительность. Мультимодальное обучение: Включение визуальной информации (например, движений губ) может дополнить аудиоданные и улучшить точность. Эффективная система ASR для каракалпакского языка может улучшить образовательные ресурсы, способствовать сохранению языка и облегчить доступ для говорящих на каракалпакском в цифровой среде. Это соответствует более широким целям языкового разнообразия и культурного сохранения во все более глобализированном мире.

Заключение. Данное исследование представляет собой комплексный подход к разработке системы ASR для каракалпакского языка с использованием модели Wav2Vec. Создавая надежный набор данных, реализуя эффективные методы предварительной обработки и аугментации, а также используя глубокое обучение, мы стремимся способствовать сохранению и цифровому представлению каракалпакского языка. Первоначальные результаты обнадеживают, прокладывая путь для будущих исследований, направленных на уточнение и улучшение системы.

Литература

1. Steffen Schneider, Alexei Baevski, Ronan Collobert, Michael Auli. Wav2Vec: Unsupervised Pre-training for Speech Recognition. 2019.

2. Hamza Kheddara, Mustapha Hemisb and Yassine Himeurc. Automatic Speech Recognition using Advanced Deep Learning Approaches: A survey. 2024.

3. Ethnologue. Languages of the World: Karakalpak. 2023.

4. Zhenyu Zhou, Shibiao Xu1, Shi Yin, Lantian Li, Dong Wang. A Com-prehensive Investigation on Speaker Augmentation for Speaker Recognition. 2024.

5. Jinyu Li. Recent Advances in End-to-End Automatic. 2020. Speech Recognition A Review of Recent Advances.

РЕЗЮМЕ. Ушбу мақола чуқур ўрганиш техникалари, хусусан, Wav2Vec моделидан фойдаланиб қароқалпоқ тили учун автоматик нутқ таниб олиш (ASR) бўйича янги ёндашувни тақдим этади. Тадқиқот микрофон орқали овоз оқимларини олиш ва ушбу аудио файлларни қайта ишлаш методологиясини баён этади.

РЕЗЮМЕ. Эта статья представляет собой новый подход к автоматическому распознаванию речи (ASR) для каракалпакского языка, использующий методы глубокого обучения, в частности модель Wav2Vec. В исследовании изложена методология захвата аудиопотоков с микрофона и обработки этих аудиофайлов.

SUMMARY. This paper presents a new approach to Automatic Speech Recognition (ASR) for the Karakalpak language, leveraging deep learning techniques, specifically the Wav2Vec model. The study outlines the methodology for capturing audio streams from a microphone, processing these audio files.