

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

УДК 004.934.1

КРЕЙС
Антон Викторович

ПРОГРАММНОЕ СРЕДСТВО РАСПОЗНАВАНИЯ РЕЧИ

АВТОРЕФЕРАТ

диссертации на соискание степени магистра

по специальности 1-40 80 04 – Информатика и технологии
программирования

Минск 2024

Работа выполнена на кафедре информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

Научный руководитель: **БОБРОВА Наталья Леонидовна,**
кандидат технических наук, доцент кафедры информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

ВВЕДЕНИЕ

Системы распознавания речи получили широкое распространение в различных сферах человеческой деятельности. В настоящее время при разработке данных систем широко используется машинное обучение, в том числе рекуррентные нейронные сети. Их архитектура хорошо подходит для работы с различными последовательностями данных, в том числе и с речевыми.

На основании обзора ряда научных работ можно сделать вывод о том, что использование рекуррентных нейронных сетей в системах распознавания речи является эффективным и перспективным.

Целью исследования настоящей магистерской диссертации является создание системы распознавания речи с использованием рекуррентных нейронных сетей, предоставляющей приемлемое качество распознавания, в виде соответствующего программного средства. Оно будет реализовано в форме текстового редактора с функцией распознавания речи. Программное средство должно предоставлять исследователю возможность сравнения эффективности использования различных конфигураций системы распознавания речи.

Для достижения поставленной цели необходимо решить ряд задач:

- изучить теоретический материал, касающийся построения систем распознавания речи;
- изучить существующие решения, используемые для распознавания речи;
- разработать перечень основных и дополнительных функций, поддерживаемых разрабатываемым программным средством распознавания речи;
- разработать архитектуру программного средства распознавания речи;
- разработать алгоритм(ы) и/или модель (модели) распознавания речи с использованием рекуррентных нейронных сетей;
- разработать графический пользовательский интерфейс программного средства распознавания речи;
- реализовать программное средство распознавания речи с использованием выбранных алгоритмов, моделей, средств разработки; сравнить между собой различные варианты использования рекуррентных нейронных сетей в программном средстве распознавания речи;
- оценить результаты выполненной работы и сделать соответствующие выводы.

Таким образом, объектом исследования настоящей магистерской диссертации являются системы распознавания речи. Предметом исследования настоящей магистерской диссертации являются различные варианты использования рекуррентных нейронных сетей в системах распознавания речи.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цели и задачи исследования

Целью исследования является создание системы распознавания речи с использованием рекуррентных нейронных сетей, предоставляющей пользователю приемлемое качество распознавания, в виде соответствующего программного средства. Для достижения поставленной цели необходимо решить ряд задач, перечень которых приведен ниже:

- изучить теоретический материал, касающийся построения систем распознавания речи;
- изучить существующие решения, используемые для распознавания речи;
- разработать перечень основных и дополнительных функций, поддерживаемых разрабатываемым программным средством распознавания речи;
- разработать архитектуру программного средства распознавания речи;
- разработать алгоритм(ы) и/или модель (модели) распознавания речи с использованием рекуррентных нейронных сетей;
- разработать графический пользовательский интерфейс программного средства распознавания речи;
- реализовать программное средство распознавания речи с использованием выбранных технологий; в процессе реализации сравнить между собой различные варианты использования рекуррентных нейронных сетей в программном средстве распознавания речи;
- оценить результаты выполненной работы и сделать соответствующие выводы.

Объектом исследования настоящей магистерской диссертации являются системы распознавания речи.

Предметом исследования настоящей магистерской диссертации являются различные варианты использования рекуррентных нейронных сетей в системах распознавания речи.

Актуальность исследования

В настоящее время системы распознавания речи занимают важное место в жизни человека. Системы распознавания речи используются в различных сферах деятельности, например:

- в различных голосовых помощниках;
- для извлечения текста из аудио- и видеозаписей;

- в телефонии для автоматизации обслуживания клиентов;
- в различном программном и аппаратном обеспечении для людей с ограниченными возможностями и т. д.

Актуальность использования систем распознавания речи состоит в том, что они позволяют автоматизировать различные рутинные задачи, а также облегчить процесс взаимодействия пользователей с разнообразным программным и аппаратным обеспечением. Последнее особенно важно для людей с проблемами со здоровьем.

Несмотря на довольно продолжительную историю систем распознавания речи, до сих пор сохраняется значительный потенциал для дальнейших исследований в данной области. Среди основных направлений исследований можно выделить повышение точности распознавания, поддержку распознавания малораспространенных языков, поддержку одновременного распознавания речи на нескольких языках, упрощение процесса обучения систем распознавания речи и др.

Вклад данного исследования в развитие систем распознавания речи состоит в получении сведений об эффективности использования рекуррентных нейронных сетей различных конфигураций при разработке систем распознавания речи. Кроме того, разрабатываемое программное средство может быть использовано для исследований эффективности использования различных конфигураций систем распознавания речи.

Публикации результатов работы

В процессе выполнения работы ее промежуточные результаты публиковались в научной литературе. Была сделана одна публикация в сборнике тезисов конференции, одна публикация в сборнике статей конференции и одна публикация в научном журнале. Общий объем публикаций составляет 10 страниц.

Личный вклад соискателя

Личный вклад соискателя состоит в сборе теоретических сведений, необходимых для проведения исследований по теме магистерской диссертации, проектировании и реализации программного средства распознавания речи, подборе параметров для модели, используемой для распознавания речи и сравнении качества распознавания при использовании различных конфигураций программного средства распознавания речи.

Структура и объем работы

Диссертация состоит из введения, общей характеристики работы, введения, трех глав, заключения, библиографического списка и приложений.

В первой главе осуществляется обзор литературы по теме магистерской диссертации. Он включает в себя сбор основных теоретических сведений по теме диссертации, обзор существующих решений распознавания речи, выбор алгоритмов и моделей для использования при разработке программного средства распознавания речи.

Во второй главе осуществляется системное проектирование программного средства. В данной главе определяется перечень его функциональных возможностей, описывается его деление на блоки, а также осуществляется выбор программных средств разработки.

В третьей главе описывается ход реализации программного средства распознавания речи. Приводится перечень файлов, составляющих программное средство, перечень и описание его классов. В данной главе приводится описание набора данных, на котором осуществляется обучение модели распознавания речи, а также описание и сравнение между собой обученных моделей. Кроме того, в данной главе приводится руководство пользователя, в котором описывается процесс взаимодействия пользователя с программным средством, а также указаны системные требования, которые соблюдались при разработке и тестировании программного средства. Глава заканчивается анализом результатов проделанной работы.

В приложении А приведена диаграмма вариантов использования программного средства распознавания речи.

В приложении Б приведен листинг исходного кода программного средства распознавания речи, а также перечень зависимостей, необходимых для работы программного средства.

Общий объем работы составляет 117 страниц, из которых основного текста – 76 страниц, 9 рисунков на 7 страницах, 1 таблица на 1 странице, библиографический список из 52 наименований на 5 страницах, включающий в себя список публикаций соискателя из 3 наименований, 2 приложения на 32 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В **введении** настоящей магистерской диссертации оценивается актуальность исследования по теме диссертации, осуществляется постановка цели диссертации и определение перечня задач, решение которых необходимо для достижения поставленной цели. Кроме того, в введении приводится общее содержание магистерской диссертации, а также определяются объект и предмет ее исследования.

В **общей характеристике работы** приводится общее описание настоящей магистерской диссертации. В ней указываются цели и задачи исследования, актуальность исследования, структура и объем работы. В общей характеристике работы также приводятся сведения о публикациях результатов работы, а также о личном вкладе соискателя.

В **первой главе** настоящей магистерской диссертации осуществляется обзор литературы по теме диссертации. Приводится определение систем распознавания речи, их классификация и история их развития. Перечисляются сферы применения систем распознавания речи и приводятся некоторые примеры данных систем. Осуществляется обзор традиционных алгоритмов и моделей, используемых при создании систем распознавания речи: DTW (Dynamic Time Warping), скрытые марковские модели (СММ). Приводятся некоторые теоретические сведения, касающиеся нейронных сетей (их определение и виды, принцип работы искусственного нейрона, виды обучения и т. д.). Приводятся теоретические сведения, касающиеся получения аудиосигнала и его предварительной обработки (этапы получения аудиосигнала, форматы его хранения на компьютере, классификация возможных шумов, присутствующих в аудиосигнале, некоторые сведения о фильтрации сигналов). Описывается принцип работы систем распознавания речи (с приведением определений некоторым терминам, связанным с распознаванием речи, описаний различных вариантов архитектур систем распознавания речи). Приводятся сведения о различных методах извлечения акустических признаков в системах распознавания речи (STFT, MFCC, LPC, вейвлет-анализ, PLP). Упоминаются некоторые разработки, которые могут быть использованы в системах распознавания речи: Connectionist Temporal Classification (CTC), RNN Transducer, Модели, основанные на внимании (Attention-based models). Перечисляются метрики качества систем распознавания речи (WER, WRR, PER, SER), а также группы факторов, влияющих на производительность систем распознавания речи. Данная глава завершается выбором моделей и алгоритмов для разработки программного средства.

Во **второй главе** настоящей магистерской диссертации осуществляется системное проектирование программного средства распознавания речи. В

данной главе описывается функционал разрабатываемого программного средства с ссылкой на диаграмму вариантов использования. Описывается деление программного средства на ряд блоков:

- получения аудиосигнала;
- извлечения акустических признаков;
- распознавания речи;
- настройки распознавания речи;
- графического пользовательского интерфейса (ГПИ);
- оценки качества распознавания речи;
- подготовки датасетов;

В данной главе приводится описание как самих блоков, так и взаимосвязей между ними.

Кроме того, в данной главе описывается выбор программных средств, используемых для разработки программного средства распознавания речи.

В **третьей** главе настоящей магистерской диссертации описывается реализация программного средства. Приводятся перечень и описание файлов программного средства. Приводятся перечень и описание классов программного средства. Описывается источник данных для обучения и тестирования программного средства распознавания речи. Описывается получение аудиосигналов, а также извлечение из них акустических признаков. Описывается процесс подбора параметров модели распознавания речи.

В рамках данной магистерской диссертации было осуществлено обучение нескольких моделей распознавания речи. Они отличаются друг от друга типом используемых в рекуррентных слоях ячеек (LSTM/GRU), их количеством. Одна из моделей обучается на разных объемах данных. Осуществляется сравнение данных моделей по различным показателям: продолжительность обучения, метрики CER и WER, размер батча и т. д.

В третьей главе размещено руководство пользователя, в котором указаны системные требования, которые соблюдались при разработке и тестировании программного средства, описываются запуск программного средства и работа с ним. Глава содержит набор изображений графического пользовательского интерфейса программного средства. Она также содержит раздел, посвященный тестированию, в котором приведен набор сценариев для E2E-тестирования.

Третья глава завершается анализом результатов работы по реализации программного средства. Также перечисляются направления по дальнейшей работе над улучшением разработанного программного средства.

ЗАКЛЮЧЕНИЕ

В результате выполнения магистерской диссертации было разработана система распознавания речи с использованием рекуррентных нейронных сетей, предоставляющая пользователю приемлемое качество распознавания, в виде соответствующего программного средства. Создание данной системы являлось целью настоящей магистерской диссертации, для достижения которой был решен ряд задач:

- был собран и изучен теоретический материал, касающийся построения систем распознавания речи;

- были изучены существующие решения, используемые для распознавания речи;

- был разработан перечень основных и дополнительных функций, поддерживаемых разрабатываемым программным средством распознавания речи;

- была разработана архитектура программного средства распознавания речи;

- были реализованы алгоритмы и модели распознавания речи с использованием рекуррентных нейронных сетей;

- был разработан графический пользовательский интерфейс программного средства распознавания речи;

- было реализовано программное средство распознавания речи с использованием выбранных алгоритмов, моделей, средств разработки; было произведено сравнение между собой различных вариантов использования рекуррентных нейронных сетей в программном средстве распознавании речи. Обученные модели показали различную эффективность распознавания речи.

- была произведена оценка результатов выполненной работы, были сделаны соответствующие выводы. Был указан ряд направлений по дальнейшему совершенствованию разработанного программного средства.

Реализованное программное средство распознавания речи предоставляет пользователю возможность самому настраивать и обучать модели и алгоритмы, используемые для распознавания речи, благодаря чему он имеет возможность сравнивать эффективность использования тех или иных моделей и алгоритмов. Таким образом, разработанное программное средство может быть использовано не только для распознавания речи, но и для разработки моделей распознавания речи (использующие метод CTC). Разработанное программное средство представляет собой пример End-to-End-системы распознавания речи.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Крейс, А. В. Системы распознавания речи / А. В. Крейс, Н. Л. Боброва // Информационные технологии и системы 2022 : материалы международной научной конференции, Минск, Беларусь, 23 ноября / редкол.: Л. Ю. Шилин (гл. ред.) [и др.]. – Минск, БГУИР, 2022. – С. 155–156.

2-А. Крейс, А. В. Извлечение акустических признаков из аудиосигнала в системах распознавания речи [Электронный ресурс] / А. В. Крейс, Н. Л. Боброва // Материалы 59-ой научной конференции аспирантов, магистрантов и студентов БГУИР по направлению «Компьютерные системы и сети» (Республика Беларусь, г. Минск, 17–21 апреля 2023 г.). – Минск, 2023. – С. 388–390. – Режим доступа: https://www.bsuir.by/m/12_100229_1_176396.pdf. – Дата доступа 25.04.2024.

3-А. Крейс, А. В. Использование рекуррентных нейронных сетей в системах распознавания речи / А. В. Крейс, Н. Л. Боброва // Студенческий вестник: научный журнал. – М. : Интернаука 2024. – № 1 (287) Ч. 9. – С. 37–41.