

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

УДК 004.934

Лежнев Антон Викторович

Алгоритмы Распознавания Речи

АВТОРЕФЕРАТ

на соискание степени магистра инженерных наук
по специальности 1-40 80 02 Системный анализ, управление и обработка
информации

Научный руководитель

Захарьев Вадим Анатольевич

доцент, кандидат технических наук

Минск 2021 г

ВВЕДЕНИЕ

Распознавание речи является одной из самых перспективных областей развития цифровых технологий. В последние годы в связи со стремительным развитием программно-аппаратных решений, алгоритмы распознавания речи сильно продвинулись в решении ряда практических задач. В настоящее время распознавание речи используется в целом ряде приложений «помощников», для построения интерфейсов человек-машина, помощи людям с ограниченными возможностями. В связи с увеличением цифровизации общества, задачами, связанными с применением распознавания речи в различных условиях, а так же необходимость снижения вычислительных мощностей необходимых для поддержки систем распознавания речи [1].

Методы распознавания речи используют различные типы входных данных, различные архитектуры для улучшения качества распознавания. В данной работе происходит исследование существующих и разработка новых алгоритмов, которые позволяют увеличить точность распознавания речи, а также снизить затраты на обучение новых моделей, и на их работу в потоковом режиме.

В первой главе диссертационной работы производится анализ предметной области, даются основные определения систем распознавания речи.

Во второй главе диссертационной работы производится анализ существующих моделей и методов распознавания речи.

В третьей главе диссертационной работы производится разработка и исследование модели на основе выбранного метода.

Исследовательская работа была опубликована в сборнике 57-ой Научной конференции аспирантов, магистрантов и студентов. Также участвовал в научной конференции согласно программе.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования

В связи со стремительно развивающимися цифровыми технологиями, а также увеличением числа мобильных устройств, набирают популярность различные приложения и системы на основе речевых технологий. На данный момент системы автоматического распознавания речи достигли значительных успехов и качества распознавания. Среди современных подходов можно выделить нейросетевой подход, который на данный момент является доминирующим в контексте разработки программных продуктов нацеленных на взаимодействие с речевым интерфейсом. Актуальность проблемы составляет несколько проблем характерных для всех нейросетевых алгоритмов и моделей-это сильная зависимость от данных, проблема переобучения и адаптация существующих архитектур для решения задач потоковой обработки речи, с минимальными вычислительными затратами.

Цель и задачи исследования

Целью данной работы является исследование алгоритмов распознавания речи, их применение на практике, разработка архитектуры на основе архитектуры Transducer с целью добиться снижения вычислительных мощностей необходимых для распознавания речи, мощностей необходимых для обучения, а также улучшения метрик распознавания.

Для выполнения поставленной цели в работе были сформулированы следующие задачи:

1. анализ существующих алгоритмов;
2. модификация существующей системы с целью улучшения характеристик распознавания;
3. анализ и сравнение с существующими системами.

Объектом исследования является система автоматического распознавания речи.

Предметом исследования являются модель преобразователя на основе которой разрабатываются нейросетевые алгоритмы распознавания речи.

Методы исследования. При решении поставленных задач использовались методы и понятия систем распознавания речи, алгоритмы и методы, применяемые в разработке нейросетевых приложений.

Научная новизна диссертационной работы заключается в исследовании новых архитектур систем распознавания речи, разработке и улучшению новых нейросетевых моделей для улучшения основных характеристик распознавания.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** приведено определение распознавания речи и актуальность темы, а также области применения рассматриваемой темы. Сформулированы цели и задачи диссертации, даны сведения об объекте и предмете исследования, указаны методы исследования.

В **первой главе** проводится анализ предметной области, даются определения основным понятиям из предметной области, рассматриваются основные моменты систем распознавания речи и приведены краткие выводы по главе.

Во **второй главе** рассматриваются методы распознавания речи такие как скрытые марковские модели, нейронные сети, дана общая структура систем распознавания. Также обосновывается выбор метода нейронных сетей, и выбор реализуемой архитектуры, происходит выбор библиотеки для реализации алгоритма на основе анализа и приводится краткий вывод по главе.

В **третьей главе** выполнено проектирование и модификация системы распознавания речи на основе архитектуры преобразователя, путём добавления свёрточных слоёв, перед слоями самовнимания в аудиокодер, представлены результаты исследования, полученные реализацией по сравнению с ближайшими аналогами по эффективности использования, представлены результаты на основе сравнения качества распознавания, показано что на основании полученных характеристик и размера аудио фрейма, данная архитектура подходит для потокового распознавания и не уступает аналогичным системам по производительности.

В **приложении** приведен листинг программного кода.

ЗАКЛЮЧЕНИЕ

Распознавание речи, позволяет приложениям, использующим определенные алгоритмы глубокого обучения, понимать гибко взаимодействовать с пользователем через интерфейс человек-машина. В процессе диссертации был разработан и исследован алгоритм распознавания речи, основанный на нейронной сети, а также проведен анализ существующих решений и методов автоматического распознавания речи.

Результаты анализа существующих решений указывают на то, что для соответствующих типов систем распознавания речи существуют свои широко и узко профильные сферы применения.

На основе анализа моделей и методов распознавания речи, был выбран метод на основе рекуррентных нейронных сетей.

В результате разработки и исследования алгоритма, основанного на нейронной сети, были выявлены его недостатки и предложены решения, данных недостатков.

Разработанный алгоритм позволяет распознавать речь с более высокой точностью, относительно других подобных методов. Самой важной проблемой в распознавании, является извлечение дискриминационных признаков и так же вычислительные мощности необходимые для выполнения обучения предложенной модели.

Для более точного и быстрого распознавания речи наряду с рекуррентными нейронными сетями, появляются решения на основе свёрточных сетей, которые позволяют за счёт модификаций частей системы значительно уменьшать затраты на обучение и работу системы, при этом не уступая аналогичным системам в качестве распознавания.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

Лежнев, А.В. Анализ современных программных средств для решения задач распознавания речи // А.В. Лежнев // Секция «Системы управления»: программа 57-й научной конференции аспирантов, магистрантов и студентов, Минск, 20 апреля 2021 г. Минск, БГУИР.

Библиотека БГУИР