

ОТЗЫВ

официального оппонента на диссертацию Пикалёва Ярослава Сергеевича на тему: «Совершенствование методов и программных средств распознавания слитной русской речи», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки).

1 Актуальность темы диссертационной работы

Технологии автоматического распознавания слитной русской речи является перспективным и динамически развивающимся направлением в области искусственного интеллекта, назначение которого заключается в обеспечении удобного диалога между диктором и машиной. Однако существующие методы распознавания голосовых команд далеки от уровня распознавания речи человеком. В широком понимании речь идет о системах, которые осуществляют фонемное декодирование речевого акустического сигнала при произношении речевых сообщений свободным стилем, произвольным диктором. Сложность, возникающая при разработке системы автоматического распознавания речи, заключается в вариативном произношении одного и того же слова как разными людьми, так и одним и тем же человеком в различных ситуациях, а также под влиянием таких факторов, как окружающий шум, отражение, эхо, помехи в канале и т.д.

В связи с этим, задача совершенствования методов и программных средств дикторонезависимого распознавания слитной русской речи, позволяющих обеспечивать высокое качество распознавания, учитывать особенности русской речи и адаптироваться под любую предметную область, является актуальной и имеет отраслевое значение.

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
Вх. № 96/189
27 109 2024

2 Обоснованность научных положений, выводов, рекомендаций, сформулированных в диссертации

Обоснованность положений, рекомендаций и выводов обеспечивается:

- полнотой анализа фундаментальных и прикладных трудов отечественных и зарубежных ученых в области распознавания слитной русской речи, что подтверждается ссылками в тексте диссертационной работы;
- корректным применением теоретических положений при построении систем дикторонезависимого распознавания слитной русской речи;
- положительной оценкой результатов работы на научных конференциях, а также в опубликованных печатных трудах;
- актом внедрения.

3 Анализ содержания диссертации

В **первой главе** выполнен анализ основных технологий автоматического распознавания русской речи, который показал, что существует необходимость использования нейросетевого подхода в распознавании речи, применяющегося в задачах акустического и языкового моделирования и классификации фонем, а также применения конечных автоматов, на основе которых появляется возможность объединить различные источники знаний. Представленная типовая архитектура систем автоматического распознавания слитной речи основана на техниках модификации извлечённых акустических признаков, цель которых заключается в повышении дикторонезависимости.

На основе выполненного анализа выделены перспективные подходы к решению задач, связанных с разработкой систем распознавания речи, а также обоснована необходимость совершенствования методов и моделей для систем дикторонезависимого распознавания слитной русской речи.

Во **второй главе** проведено исследование методов построения акустической и языковой моделей для создания и обучения речевого корпуса.

Выполнено описание алгоритма проверки соответствия текстовых и аудио расшифровок, позволяющего автоматизировать процесс создания русскоязычного речевого корпуса, а также приведено описание методов аугментации, нормализации русскоязычного текста, используемых для обучения языковых моделей и систем автоматического исправления ошибок.

В **третьей главе** рассмотрены особенности фонетики русского языка и предложена система автоматического получения транскрипций, приведена общая схема функционирования этой системы, а также архитектура нейросети, используемая для определения положения ударения, генерации транскрипций для слов-исключений и практической транслитерации. Система автоматической генерации транскрипций основана на гибридном подходе, который использует словари и нейросети. Анализ результатов моделирования показал, что предложенные модификации позволяют повысить точность генерации транскрипции по сравнению со стандартными подходами.

В **четвертой главе** предложена технология повышения робастности акустической модели в задаче распознавания речи с применением глубокого и машинного обучения. Предлагаемая технология основана на использовании информативных акустических признаков, извлечённых из иерархических нейросетевых моделей, а также на гибридных акустических моделях, обученных на основе машинного и глубокого обучения с применением дискриминативного подхода.

В **пятой главе** представлена структура разработанной системы автоматического распознавания слитной русской речи, рассмотрены основные её структурные блоки, описан процесс получения языковой и акустической модели, процесс декодирования. Охарактеризованы обучающие и тестовые корпуса текстовых и речевых данных, проведено сравнение эффективности работы разработанной системы с существующими решениями.

4 Научная новизна работы

Научная новизна полученных результатов заключается в следующем:

– получили дальнейшее развитие нейросетевые методы автоматического определения позиции ударения в слове за счет модернизации архитектуры нейросети типа Transformer, которая заключается в увеличении количества слоёв, использовании методов градиентного отсечения и teacher forcing для оптимизации параметра скорости обучения, что позволило повысить точность определения позиции ударения на 10% по сравнению со стандартной моделью Transformer;

– усовершенствована seq2seq модель для генерации практических транскрипций англоязычных слов и слов-исключений за счет применения механизма обучения с подкреплением и метода beam-search для выбора наиболее вероятной последовательности символов, что позволило повысить точность модели по критерию количества ошибочно сгенерированных символов на 0,8% и 3%, по критерию неправильно сгенерированных слов на 0,6% и 9% соответственно;

– предложена модель нейросетевой параметризации, основанная на объединении ансамбля нейронных сетей с «узким горлом» и архитектуры ResNet-50, использование которой позволяет повысить точность распознавания на 2,7% по сравнению с моделью, извлекающей стандартные bottleneck-признаки;

– получили дальнейшее развитие методы нейросетевой классификации фонем за счет использования механизма внимания в последнем скрытом слое сети, включающей в себя нейросеть с временными задержками и двунаправленную нейросеть с долгой кратковременной памятью, что позволило сохранять высокую точность на относительно небольшом обучающем наборе аудиоданных, свойственную системам, для обучения которых требуется речевая база длительностью в десятки тысяч часов.

5 Практическое значение полученных результатов

Материалы исследований могут быть использованы при разработке методов автоматического формирования аннотированных речевых баз данных, методов автоматического построения словаря транскрипций в системах синтеза речи,

методов получения робастных акустических признаков и обучения АМ, а также при разработке классификатора для распознавания фонем в системах голосового управления и поиска по голосовому запросу, а также в системах диктовки с приемлемым уровнем ошибок.

Результаты и выводы диссертационной работы нашли применение в Институте проблем искусственного интеллекта, что подтверждается справкой о внедрении (справка №347/01-01 от 01.12.2020).

6 Степень полноты опубликованных полученных результатов

Основные научные результаты диссертации опубликованы в 17 печатных работах, в том числе в 5 научных статьях в изданиях, рекомендуемых ВАК для публикации трудов на соискание ученых степеней.

7 Соответствие содержания диссертации специальности 05.13.01 – Системный анализ управление и обработка информации (по отраслям) (технические науки)

Содержание рецензируемой диссертационной работы, выдвинутые научные положения, полученные выводы и рекомендации дают основание сделать заключение о том, что диссертация Пикалёва Я. С., представленная к защите на соискание ученой степени кандидата технических наук, соответствует паспорту специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки).

8 Соответствие автореферата содержанию диссертации

В автореферате в достаточном объеме изложены основные результаты исследований, приведенных в диссертации. В нем представлены научные положения, новизна и практическая значимость, основные выводы по 4 главам

диссертации и работе в целом, заключение, информация о реализации результатов работы и предложения по их дальнейшему использованию.

9 Структура, стиль и язык диссертации

Структурное построение диссертации соответствует цели и задачам исследований. Стиль изложения содержания исследований и подача материала логичны, последовательны и связаны единой идеей. Язык диссертации достаточно ясен и доступен для восприятия.

10 Замечания по диссертации

1. В главе 2 стр. 26 указано: «Обучение останавливается, как только абсолютное уменьшение ошибки классификации кадров окажется достаточно малым». Из текста диссертации не ясно, каким образом получено значение «достаточно малым» и что под ним подразумевается.

2. В главе 2 стр. 60 указано: «В качестве метрик использовались стандартные метрики потерь и точности классификации (рисунок 2.4), а также F-мера (F1-score) (рисунок 2.5). Проведенные исследования показали, что точность модели определения языка текста составила на тестовой выборке 82.5%, на обучающей – более 95%». Из текста диссертации не ясно, по какой формуле определяется F-мера и каким образом значения точности модели появляются, так как в тексте диссертации они не представлены.

3. В главе 4 стр. 130 указано: «Для получения робастных акустических признаков и обучения АМ предложена нейросетевая параметризация, основанная на объединении ансамбля нейронных сетей с узким горлом и архитектуры ResNet-50, что позволяет повысить точность распознавания на 2,7% по сравнению с моделью, извлекающей стандартные bottleneck-признаки;». Из диссертации не ясно, каким образом данные значения точности распознавания появляются в заключении, так как в тексте диссертации они не представлены.

4. В главе 5 стр. 151 указано: «Разработана нейросетевая модель денормализации распознанного текста на базе архитектуры Transformer, позволяющая восстанавливать знаки препинания, строчные числа заменять цифрами, восстанавливать прописные буквы в именах собственных и аббревиатурах, которая обеспечивает точность 0,87.». Из диссертации не ясно, каким образом данные значения точности появляются в заключении, так как в тексте диссертации они не представлены.

5. В заключении п. 2 указано «В результате использования предложенных алгоритмов для обучения АМ подготовлен речевой корпус длительностью более 29 часов, для обучения ЯМ сформирована база нормализованных текстов объемом 15,2 Гб...», а в п. 7 – «На основе предложенных методов и моделей разработана ASR-система, которая обучалась на речевом корпусе объемом около 7,2 Гб.». Какой же объём данных был использован?

6. В диссертационной работе не указано, какой именно программный комплекс разработан для системы автоматического распознавания речи.

11 Заключение

Диссертационная работа Пикалёва Ярослава Сергеевича является законченной научно-квалификационной работой, в которой дано теоретическое обоснование и приведено решение важной научно-практической задачи совершенствования методов и программных средств распознавания слитной русской речи для дикторонезависимых систем, позволяющих обеспечивать высокое качество распознавания, учитывать особенности русской речи и адаптироваться под любую предметную область.

В целом, работа выполнена на достаточно высоком научном уровне, содержит новые теоретические и полезные практические результаты. Отмеченные недостатки не препятствуют положительной оценке работы. Содержание диссертации, положения и выводы соответствуют специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям)

(технические науки), а сама работа соответствует требованиям ВАК Донецкой Народной Республики, предъявляемым к кандидатским диссертациям и удовлетворяет требованиям п. 2.2 «Положения о присуждении ученых степеней», утвержденного Постановлением Совета Министров Донецкой Народной Республики от 27.02.2015 г. №2-13.

Работа заслуживает положительной оценки, а соискатель Пикалёв Ярослав Сергеевич заслуживает присуждения степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки).

Официальный оппонент,

Инженер по автоматизированным системам управления производством
отдела сетевых сервисов дирекции по информационным технологиям
ГУП ДНР «ЭНЕРГИЯ ДОНБАССА»,

кандидат технических наук по специальности
05.13.01 – Системный анализ, управление и
обработка информации (по отраслям)

(технические науки)

Адрес: 283114, г. Донецк, ул. Щорса, 87,

тел.: (071) 314-55-05

эл.почта: delo@don.energy

сайт: <https://don.energy/>

Е. И. Бурлаева

Я, Бурлаева Екатерина Игоревна, даю согласие на автоматизированную обработку моих персональных данных.

Кандидат технических наук

Е. И. Бурлаева

Подпись Бурлаевой Екатерины Игоревны подтверждаю



И. С. Скитер