

ОТЗЫВ

на автореферат диссертации

Пикалёва Ярослава Сергеевича

на тему «**Совершенствование методов и программных средств
распознавания слитной русской речи**»,

представленной на соискание ученой степени кандидата технических наук
по специальности 05.13.01 – Системный анализ, управление и обработка
информации (по отраслям) (технические науки)

Актуальность темы диссертации обусловлена необходимостью совершенствования методов и моделей распознавания русской речи с целью повышения точности распознавания в условиях шумов, вызванных различными каналами связи, а также окружающей средой. При разработке русскоязычных систем распознавания слитной речи стоит выделить ряд фонетических проблем, связанных с редукцией гласных, ослаблением и выпадением согласных, уменьшением степени контрастности между гласными и согласными в пределах слога, которые увеличивают акустическую вариативность для русскоязычной слитной речи. При статистическом моделировании русского языка и получении качественной языковой модели возникают трудности, связанные с его флективностью и свободном порядком слов в предложении.

Диссертационное исследование Пикалёва Ярослава Сергеевича посвящено разработке методов и алгоритмов распознавания слитной русской речи, учитывающих её фонетические и лингвистические особенности и повышающих качество распознавания.

В представленной диссертации получили развитие нейросетевые методы автоматического построения транскрипций слов русского языка и практической транслитерации 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;
- применения механизма обучения с подкреплением и метода beam-search для выбора наиболее вероятной последовательности символов к

seq2seq модели для генерации практических транскрипций англоязычных слов и слов-исключений, 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.

Кроме того, автором предложены модели:

- нейросетевой параметризации, основанная на объединении ансамбля нейронных сетей с «узким горлом» и архитектуры ResNet-50. Использование данной модели позволяет повысить точность распознавания на 2,7% по сравнению с моделью, извлекающей стандартные bottleneck-признаки;

- нейросетевой классификации фонем, использующие механизм внимания в последнем скрытом слое сети, которая базируется на архитектурах сети с временными задержками и двунаправленной сети 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.

Анализ содержания автореферата позволяет сделать заключение, что диссертация Пикалёва Ярослава Сергеевича является самостоятельно выполненной научно-квалифицированной работой.

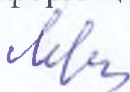
Также следует отметить, что разработанные техники обработки данных для обучения акустических и языковых моделей быть использованы при разработке методов автоматического формирования аннотированных речевых баз данных; предложенные алгоритмы формирования словарей транскрипций найдут применение в системах синтеза речи; разработанные методы получения робастных акустических признаков и классификации фонем будут полезны при разработке систем голосового управления и поиска по голосовому запросу, а также в системах диктовки с приемлемым уровнем ошибок.

Вместе с тем имеются некоторые замечания.

Для исправления текстовых ошибок автором используется языковая модель ELMo, которая генерирует векторное представление для слова на основе его контекста. Модель улучшена при помощи техник teacher forcing и

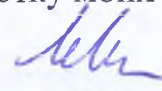
градиентного отсечения, за счет чего предложенная архитектура работает быстрее, чем стандартно используемые в NLP-задачах LSTM. Однако при разработке системы распознавания на декодер поступают вероятности, генерируемые стандартной языковой моделью на базе LSTM. Неясным остается выбор автора в пользу стандартной архитектуры LSTM при декодировании, а не улучшенной модели на базе ELMo.

В целом, несмотря на отмеченные замечания, диссертация Пикалёва Ярослава Сергеевича является завершённой научно-исследовательской работой. По структуре, содержанию и объёму диссертационная работа отвечает требованиям, предъявляемым к кандидатским диссертациям, в соответствии с Положением о присуждении ученых степеней, а соискатель, Пикалёв Ярослав Сергеевич, заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки).

Доктор технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (по отраслям) (технические науки)»,
Профессор  Виталий Ильич Левин

Адрес: 440039, г. Пенза, проезд Байдукова/ул. Гагарина, 1а/11
тел.: +7(986)9421507
E-mail: vilevin@mail.ru

Федеральное государственное бюджетное образовательное учреждение высшего образования «Пензенский государственный технологический университет» (ПензГТУ), ведущий научный сотрудник

Согласен на автоматизированную обработку моих персональных данных
 Виталий Ильич Левин

Подпись Левина Виталия Ильича подтверждаю:

