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)

Речевые технологии предоставляют возможность общения компьютера и человека посредством речи без использования каких-либо иных способов ввода информации и имеют широкую сферу применения: в системах диктовки, компьютерной телефонии, интерфейсах для пользователей-инвалидов, слепых или слабовидящих, диспетчерских системах управления, обучения иностранным языкам и т.д.

В связи с этим наиболее актуальным и востребованным видом интеллектуальных систем являются автоматизированные системы распознавания речи.

Однако, системы распознавания русской речи в связи с ее высокой изменчивостью пока еще не обладают той степенью достоверности распознавания, которая требуется для их практического применения. Дикторонезависимое распознавание слитной речи осложняется некоторым количеством посторонних звуков, создаваемых диктором, внедикторскими вариациями, слова часто проговариваются небрежно, говорящий может находиться в сложной акустической обстановке, все это приводит к снижению точности распознавания.

Указанные обстоятельства свидетельствуют об актуальности разработки методов и программных средств распознавания слитной русской речи. Поэтому тема данной диссертации, без сомнения, является **актуальной**.

Основные научные положения, заявленные в автореферате и вынесенные на защиту, соответствуют поставленным целям и задачам работы, а также коррелируют с объектом и предметом исследования. Результаты исследования отражены в 17 научных публикациях и прошли апробацию на многих международных конференциях. Реализация выводов и рекомендаций работы подтверждается внедрением результатов исследований диссертационной работы в Институте проблем искусственного интеллекта при выполнении научно-исследовательских работ (справка №347/01-01 от 01.12.2020).

Полученные результаты, положения и выводы отвечают требованиям пунктов 5 и 12 паспорта специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки).

При выполнении работы автором использован современный математический аппарат: в работе приведены выводы и заключения, основанные на методах цифровой обработки сигналов, методах машинного обучения и методах математической статистики.

Автором рассмотрены современные технологии распознавания слитной русской речи, на основе которых выбраны методы построения акустической и языковой модели, транскриптора, классификатора фонем и декодера. Для обучения моделей и создания аннотированного речевого корпуса собраны и обработаны речевые и текстовые данные, находящиеся в открытом доступе. Для создания транскриптора разработаны методы автоматического построения словаря транскрипций, позволяющие определять позицию ударения, генерировать транскрипцию для слов-исключений и осуществлять практическую транслитерацию с учетом орфоэпических норм русского языка. Для создания акустической модели разработаны методы получения робастных акустических признаков, а также классификатор для распознавания фонем на основе глубоких нейронных сетей. Предложенные методы и модели реализованы в единой системе распознавания слитной русской речи и проведена оценка качества ее работы по сравнению с российскими и зарубежными аналогами. Численные исследования показали эффективность разработанных методов. Авторская система, обладая достаточной точностью в задаче распознавания слитной русской речи, использует для своего обучения значительно меньше данных и превосходит рассмотренные системы по скорости получения результата распознавания более, чем в 7 раз.

При анализе автореферата, представленного на рецензию, считаю важным отметить наиболее значимые положения научной новизны:

1) для автоматического определения позиции ударения в слове получили дальнейшее развитие нейросетевые методы за счет модернизации архитектуры нейросети типа Transformer путем увеличения количества слоёв, использования методов градиентного отсечения и teacher forcing, оптимизирующего скорость обучения. Предложенная модификация позволила повысить точность определения позиции ударения на 10% по сравнению со стандартной моделью Transformer;

2) для генерации практических транскрипций англоязычных слов и слов-исключений усовершенствована seq2seq модель путем 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;

3) для получения высокоуровневых акустических признаков предложена модель нейросетевой параметризации, основанная на объединении ансамбля нейронных сетей с «узким горлом» и архитектуры ResNet-50. Предложенный подход к построению акустической модели позволяет повысить точность распознавания на 2,7% по сравнению с моделью, извлекающей стандартные bottleneck-признаки;

4) для нейросетевой классификации фонем предложена архитектура, включающая в себя нейросеть с временными задержками и двунаправленную нейросеть с долгой кратковременной памятью, в последнем скрытом слое сети данной архитектуры используется механизм внимания. Такая архитектура позволяет сохранять высокую точность на относительно небольшом обучающем наборе аудиоданных, свойственную системам, для обучения которых требуется речевая база длительностью в десятки тысяч часов.

Автореферат дает полное представление о содержании работы. Вместе с тем имеется ряд замечаний:

1) для адаптации акустической модели под диктора используются i-вектора, которые содержат суммарную информацию о голосе диктора, помехах и искажении сигнала в канале связи, однако нет количественных данных о допустимом уровне помех в речевом сигнале. При невысоком отношении сигнал-шум качество распознавания сильно падает. Было бы целесообразным при обучении акустической модели отдельно учесть влияние помех.

2) для увеличения объема обучающих данных и повышения робастности акустической модели предложена техника аугментации речевых данных путем наложения шумов и голоса другого диктора. Речевой шум приводит существенно снижает точность распознавания. Следовало бы провести дополнительные исследования влияния уровня речевого шума на качество акустической модели.

В целом указанные замечания не влияют на высокую оценку диссертации, представляющей собой завершенную научно-исследовательскую работу. По своей значимости, научной новизне и практическому значению данная работа в полной мере соответствует действующим требованиям к кандидатским диссертациям и паспорту специальности 05.13.01 – Системный анализ, управление и обработка

информации (по отраслям) (технические науки). На основании вышесказанного полагаю, что Пикалёв Ярослав Сергеевич заслуживает присуждения ученой степени кандидата технических наук по заявленной специальности.

Кандидат технических наук по специальности
05.13.06 – Автоматизация и управление
технологическими процессами и производствами
(по отраслям) (технические науки),
ст. преподаватель кафедры «АТС и ВТ»
ГОСУДАРСТВЕННОЙ ОБРАЗОВАТЕЛЬНОЙ
ОРГАНИЗАЦИИ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ДОНЕЦКИЙ ИНСТИТУТ
ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА»

_____ Трунаев Андрей Михайлович

Адрес: 283018, г. Донецк, ул. Горная, 6
тел.: +38 (071) 331-25-70
E-mail: Davidoff-a@mail.ru

Согласен(сна) на автоматизированную обработку моих персональных данных

_____ Трунаев Андрей Михайлович

Подпись Трунаева Андрея Михайловича подтверждаю:

НАЧАЛЬНИК ОТДЕЛА КАДРОВ
ДОНЕЦКОГО ИНСТИТУТА
ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА

ОТДЕЛ КАДРОВ

В. Н. Гончарук