Разработка и исследование интеллектуальной системы для комплексного паралингвистического анализа речи

Аннотация результатов, полученных в 2018 году

Выполненные на первом этапе проекта в 2018 отчетном году работы включают в себя расширенный аналитический обзор существующего информационно-лингвистического и математического обеспечения, связанного с областью компьютерной паралингвистики, разработку новых и совершенствование существующих моделей, методов и алгоритмов комплексного паралингвистического анализа речи, а также сбор и анализ доступных многодикторных речевых корпусов (информационно-лингвистическое обеспечение) на различных естественных языках для многоцелевых исследований паралингвистических речевых явлений.

1) Аналитический обзор предметной области включает в себя более 140 научно-технических источников литературы, более 130 из которых были опубликованы за последние 7 лет. В обзор вошли такие разделы, как описание предметной области; актуальные цели и задачи компьютерной паралингвистики; способы сбора, аннотации и обработки существующих речевых корпусов на разных языках, представляющие различные паралингвистические явления, в том числе психоэмоциональные состояния, ложь, стресс, состояние здоровья, сна, а также индивидуальные характеристики дикторов, такие, как пол, возраст, языковой акцент и другие. В обзор также вошло описание открытых международных соревнований по компьютерной паралингвистике ComParE, ежегодно проходящих в рамках международной конференции INTERSPEECH, в которых участвовали исполнители данного проекта. Помимо этого, в обзоре были описаны современные достижения и передовые технологии, применяющиеся в области автоматического анализа паралингвистических явлений: представлены методы извлечения и обработки акустических признаков, их нормализации; поиск оптимальных представлений информативных признаков; алгоритмы

классификации, включающие в себя как традиционные методы машинного обучения, так и самые современные нейросетевые архитектуры; способы обучения классификаторов, в том числе, использование кросс-корпусного анализа и аугментации данных. Дана классификация существующих и применяющихся на практике методов машинного обучения в области компьютерной паралингвистики, проанализированы преимущества и недостатки каждого метода, приведены примеры успешного применения подходов на практике. Также приведена классификация типов акустических признаков, играющих важную роль в системах распознавания паралингвистических событий, даны основные характеристики каждой группы.

2) После поиска и сбора существующего открытого информационно-лингвистического обеспечения были проанализированы следующие базы данных, содержащие различные паралингвистические явления: базы данных эмоционально окрашенной русской речи: RUSLANA, RAMAS, EmoChildRu; базы данных эмоциональной речи на других языках: англоязычные – IEMOCAP, CreativeIT, SEMAINE; немецкоязычные – EMODB, USoM; франкоязычная – RECOLA; турецкоязычная – BUEMODB. Базы данных с правдивыми и ложными речевыми сообщениями: Deceptive Speech Database (DSD), CSC Deceptive Speech (CSC), база данных Университета Ноттингема, база данных Университета Сучжоу (Китай), корпус Columbia X-Cultural Deception Corpus (CXD Corpus). Базы данных, содержащие индивидуальные характеристики дикторов (пол и возраст): aGender, ELSDSR, Mandarin, NIST SRE 2008, NIST SRE 2010, N-Best. Другие найденные речевые базы данных с паралингвистическими явлениями: базы данных, содержащие речь людей с болезнью Паркинсона; базы данных, содержащие речь с различными акцентами. В общей сложности для проведения исследований было получено более 20 свободно-доступных речевых корпусов, содержащих различные паралингвистические явления. Из анализа полученных открытых баз данных можно сделать вывод, что задача распознавания эмоций превалирует в

современной области компьютерной паралингвистики, что видно из большого количества существующих данных. Корпуса лживой речи значительно уступают по количеству и объему, а также возможности открытого использования данных в целях исследований. Для распознавания возраста и пола подходят практически любые базы данных, содержащие базовую информацию о дикторах. Одной из самых малоисследованных задач в области компьютерной паралингвистики является распознавание наличие заболеваний по речи (например, болезни Паркинсона или Альцгеймера), в силу малого количества подходящих дикторов, а также конфиденциальности процедуры записи пациентов.

3) В целях разработки интеллектуальной системы комплексного паралингвистического анализа речи было предложено новое и усовершенствовано существующее математическое обеспечение для вычисления и выделения акустических признаков, в том числе:

а) Усовершенствован метод извлечения акустических признаков с использованием программного инструментария openSMILE, который предоставляет возможность извлечения стандартных акустических признаков на уровне всего высказывания. Количество таких признаков очень высоко (более 6 тыс.), поэтому существует необходимость предварительной обработки векторов признаков перед их использованием для классификации. Нами предложен новый подход к нормализации полученных признаков с помощью каскадного применения операций нормализации на разных уровнях цифровой обработки аудиосигнала. В результате такой метод обработки позволяет избавиться от вариативности между дикторами, сократить диапазон изменения значений различных признаков и привести любой набор данных к удобному виду для обработки, что приводит к повышению эффективности работы классификаторов.

б) Предложен метод извлечения признаков на каждом кадре при помощи рекуррентной нейронной сети с длинной кратковременной памятью (РНС-ДКП), которая известна своей эффективностью при моделировании

временных последовательностей. Специальные ячейки памяти ДКП позволяют хранить информацию о предыдущих событиях, закодированную в виде активации соответствующих параметров сети. Такая архитектура имеет преимущество перед обычными РНС, которое выражается, во-первых, в возможности моделировать неограниченно длинные последовательности, и, во-вторых, в отсутствии проблемы взрывающихся градиентов, присущей архитектурам РНС без ДКП. Предложенный способ извлечения признаков позволяет моделировать временные изменения в речевом сигнале, захватывать динамическую структуру данных и генерировать абстрактное представление, содержащее скрытые информативные признаки, недоступные для выражения через стандартные статистические методы openSMILE. Таким образом, комбинация глобальных статистических признаков openSMILE и признаков, полученных на выходе РНС-ДКП, моделирующих локальную структуру данных и ее изменения во времени, позволяет воспользоваться преимуществами обоих методов, которые являются комплиментарными.

в) Предложен новый подход к кросс-корпусному обучению классификаторов, позволяющий использовать больше данных для обучения и добиться более высокой точности и робастности классификации. В ходе данной работы было сделано несколько выводов, в том числе о том, что измерения активации и валентности являются сильно коррелированными, поэтому обучение системы на активации эмоции, а тестирование на валентности эмоции, показывает высокие результаты. Это верно в том случае, если исходное распределение эмоциональных дескрипторов в пространстве активация-валентность обучающего корпуса имеет положительную корреляцию, как например, для корпусов RECOLA и SEMAINE. По этой причине при использовании кросс-корпусного обучения необходимо тщательно подбирать корпуса, подходящие по характеристикам распределения к целевому корпусу тестирования.

г) Предложен новый метод к генерации новых данных на основе уже имеющихся обучающих наборов из разных корпусов. Данный метод является простой, но эффективной стратегией генерации новых данных для преодоления сложностей, связанных с кросс-корпусным моделированием на несовпадающих распределениях обучающих и целевых ковариационных структур.

4) По результатам выполненных работ были подготовлены и опубликованы 3 статьи в англоязычных изданиях (Lecture Notes in Computer Science и Proceedings of INTERSPEECH), индексируемых в базах данных Scopus/Web of Science и 2 статьи в изданиях, индексируемых в РИНЦ, подготовлена 1 журнальная статья, которая принята к публикации в 2019 г. Сделаны устные доклады по текущим результатам проекта на международных научных конференциях INTERSPEECH-2018 (Хайдерабад, Индия, 2-6 сентября 2018), SPECOM-2018 (Лейпциг, Германия, 18-22 сентября 2018) и ИТУ/МКПУ-2018 (Санкт-Петербург, 4-6 октября 2018). Кроме того, авторы приняли участие в 10-м паралингвистическом соревновании ComParE в рамках 19-й международной конференции INTERSPEECH-2018 по конкурсному направлению определения оценки эмоционального состояния дикторов, предоставленной самими дикторами в виде самооценки (Self-Assessed Sub-Challenge), и оказались в числе 3-х лучших финалистов конкурса. В целях освещения результатов данного проекта создана веб-страница в глобальной сети Интернет, посвященная данному исследованию: http://hci.nw.ru/ru/projects/18

Публикации

1. *Величко А.Н., Карпов А.А., Будков В.Ю.* **Аналитический обзор речевых корпусов для систем определения ложных речевых сообщений** Материалы конференции «Информационные технологии в управлении» (ИТУ-2018), Санкт-Петербург, ИТУ-2018, С. 534-538 (год публикации - 2018).

2. *Верхоляк О.В., Кайя Х., Карпов А.А.* **Modeling short-term and long-term dependencies of the speech signal for paralinguistic emotion classification** Труды СПИИРАН (SPIIRAS Proceedings), - (год публикации - 2019).

3. *Кайа Х., Федотов Д., Йешилканат А., Верхоляк О.В., Жанг Й., Карпов А.А.* **LSTM Based Cross-corpus and Cross-task Acoustic Emotion Recognition** Proceedings of the Annual Conference of the International Speech Communication Association INTERSPEECH, INTERSPEECH-2018, с. 521-525 (год публикации - 2018).

4. *Маркитантов М.В., Карпов А.А.* **Аналитический обзор подходов к автоматическому распознаванию возраста диктора по голосу** Материалы конференции «Информационные технологии в управлении» (ИТУ-2018), Санкт-Петербург, ИТУ-2018, С. 539-542 (год публикации - 2018).

5. *Марковников Н.М., Кипяткова И.С., Лякso Е.Е.* **End-to-End Speech Recognition in Russian** Lecture Notes in Computer Science, т. LNAI 11096, с. 377–386 (год публикации - 2018).

6. *Федотов Д., Кайа Х., Карпов А.А.* **Context Modeling for Cross-Corpus Dimensional Acoustic Emotion Recognition: Challenges and Mixup** Lecture Notes in Computer Science, т. LNAI 11096, с. 155–165 (год публикации - 2018).