

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

УДК 159.942:004.522

Брановицкий
Артем Александрович

Модели и методы распознавания эмоций на основе машинного обучения

АВТОРЕФЕРАТ

на соискание академической степени
магистра технических наук

по специальности 1-40 80 05 – Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Научный руководитель
Серебряная Л.В.
к. т. н., доцент

Минск 2020

ВВЕДЕНИЕ

В настоящее время во всём мире активизировались исследования в области искусственного интеллекта, в том числе эмоционального ИИ. Им интересуются крупные компании и амбициозные стартапы, которые внедряют новые разработки во многие сферы человеческой жизни. По данным Mordor Intelligence, рынок распознавания эмоций оценивался в \$12 млрд в 2018 году и вырастет до \$92 млрд к 2024 году.

Эмоциональный искусственный интеллект – это ИИ, который позволяет компьютеру распознавать и интерпретировать человеческие эмоции и реагировать на них. Камера, микрофон или нателный датчик считывают состояние человека, а нейросеть обрабатывает данные, чтобы определить эмоциональное состояние человека.

Технологии распознавания эмоций человека нашли широкое применение в разнообразных областях, перечислим часть из них:

1. Продажи и реклама. С помощью распознавания эмоций определяется, как рекламный ролик действует на человека. Для этого можно, например, установить конструкцию с камерой, которая будет менять рекламу в зависимости от настроения, пола и возраста проходящих мимо людей.

2. Банковская сфера. Технологии распознавания эмоций помогают банкам получать обратную связь от клиентов без опросов и улучшать обслуживание. В отделениях устанавливают видеокamеры, и алгоритмы по записи определяют удовлетворенность посетителей. Нейросети также используются для анализа голоса и речи клиента и оператора во время звонка в контакт-центр.

3. Сфера развлечений. Системы распознавания эмоций используются для того, чтобы узнать реакцию аудитории на фильм. Также распознавание эмоций уже используется для создания интерактивных интерфейсов – игр следующего поколения, где сюжет может меняться в зависимости от эмоций игрока.

4. Автомобили. Массачусетский технологический институт разрабатывает ИИ под названием AutoEmotive, который определит состояние водителя и пассажиров. Он будет не только следить за уровнем стресса, но и пытаться снизить его – включать спокойную музыку, регулировать температуру в салоне или выбирать менее загруженный маршрут.

В сложившейся ситуации создание инструмента для распознавания эмоционального состояния человека представляется актуальной задачей. Одну из главных ролей в данном инструменте играет машинное обучение.

Машинное обучение – это инструмент, за последние годы получивший широкое применение в программных средствах разного масштаба и направленности – от сложнейших комплексов программного прогнозирования до развлекательных мобильных приложений. Его ценность в возможности создания динамических систем, перестраивающихся в соответствии с изменениями в предметной области.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является исследование моделей и методов распознавания эмоций и проектирование эффективной системы распознавания эмоций по голосу на основе машинного обучения.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Изучить научно-методическую и справочную литературу по вопросу распознавания эмоций человека.
2. Проанализировать существующие методы выделения значимых признаков (характерных особенностей) речевых сигналов.
3. Выполнить поиск и анализ моделей и методов классификации в сфере распознавания эмоций.
4. Сравнить эффективность классических алгоритмов машинного обучения (DTA, SVM и др.) и нейронных сетей различных архитектур (BPNN и её модификаций, PNN и др.) для решения задач распознавания эмоций.
5. Разработать методы по увеличению количества тестовых данных в связи с тем, что БД, содержащие речевые файлы, имеют малый размер.
6. Спроектировать систему распознавания эмоций по голосу на основе наиболее эффективных методов выделения и выбора значимых признаков речи и моделей классификации, выявленных в ходе сравнительного анализа.

Объектом исследования является эмоциональное состояние человека.

Предметами исследования являются методы выделения характерных особенностей из речевых сигналов и модели классификации эмоциональных состояний.

В основу данной диссертационной работы положены следующие *гипотезы*:

1. Речь человека передаёт его эмоциональное состояние.
2. Существует возможность вычисления акустических параметров речи, характеризующих эмоциональное состояние человека.
3. Существует модель классификации, применимая к имеющимся акустическим параметрам речи и способная производить достаточно точную классификацию эмоциональных состояний.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Программное обеспечение информационных технологий» по теме «Разработка моделей, методов, алгоритмов, повышающих показатели проектирования, внедрения и эксплуатации программных средств для перспективных платформ обработки информации, решения интеллектуальных задач, работы с большими массивами данных и внедрение в современные обучающие комплексы» (ГБ № 16-2004, № ГР 20163588, научный руководитель НИР — Н. В. Лапицкая).

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Л. В. Серебряной, заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на международной научной конференции «Информационные технологии и системы» (Минск, Беларусь, 30 октября 2019).

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 печатных работы, из них 1 работа в сборниках трудов и материалов международной конференции, 1 работа в сборниках трудов и материалов внутривузовской конференции.

Структура и объем диссертации

Диссертация состоит из общей характеристики работы, введения, пяти глав, заключения, списка использованных источников, списка публикаций автора и приложения. В первой главе представлен обзор предметной области, выявлены основные существующие подходы, методы и алгоритмы в рамках тематики исследования. Во второй главе рассматривается проблематика классификации эмоциональных состояний человека. В ней приводится описание машинного обучения как методики, рассматриваются классические модели машинного обучения и нейросетевые модели. Третья глава рассматривает проблематику вычисления значимых признаков речевых сигналов. В четвёртой главе формулируются функциональные требования к разрабатываемому ПС, приводятся результаты проектирования архитектуры разрабатываемого ПС, осуществляется выбор базы данных для классификатора, архитектуры МП, языка программирования и набора прикладных библиотек. В пятой главе производится поиск оптимальной конфигурации спроектированной системы распознавания эмоций по речевому сигналу на основании выбора различных подмножеств значимых признаков речевого сигнала для расчёта и модификации реализованных моделей классификации. Дается оценка эффективности созданной системы распознавания эмоций по речевому сигналу в оптимальной конфигурации в сравнении с уже существующими системами.

Общий объем работы составляет 77 страниц, из которых основного текста — 58 страниц, 37 рисунков на 32 страницах, 8 таблиц на 7 страницах, список использованных источников из 33 наименований на 3 страницах и 1 приложение на 10 страницах.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Данная диссертационная работа посвящена моделям и методам распознавания эмоций на основе машинного обучения.

В первой главе рассматривается проблематика категоризации эмоций. Суть которой заключается в том, что эмоции субъективны, и даже люди интерпретируют их по-разному. Приводится обзор категориального, размерного и оценочного подходов для классификации эмоций. Производится обзор классического (извлечение ключевых точек и пространственных дескрипторов с подачей на вход классификатору) и основанного на глубоком обучении (изображение подаётся на вход нейронной сети, содержащей свёрточные и рекуррентные слои) подходов к распознаванию эмоций по выражению лица. Приводится обоснование выбора речевых моделей для распознавания эмоций в качестве предмета исследования в связи с тем, что у них есть много потенциальных областей для применения в системах взаимодействия человек-компьютер, человек-человек, к тому же речь является второй по значимости характеристикой (после мимики) при восприятии эмоций человеком и проблема автоматического распознавания эмоционального состояния говорящего по голосу на данный момент не является решённой. Производится обзор и анализ программ-прототипов. Рассматривается структура системы распознавания эмоций по речи.

Во второй главе приводятся общие сведения о машинном обучении, целесообразности его применения для решения тех или иных задач. Делается вывод об оправданности его использования для создания системы распознавания эмоций по речевому сигналу, так как сложно заранее сформулировать формальные правила, по которым можно определить эмоциональное состояние человека по речевому сигналу. Производится выбор метода опорных векторов и деревьев решений как представителей классических моделей машинного в качестве классификаторов для системы распознавания эмоций по речевому сигналу, потому что простые модели машинного обучения более просты в понимании и интерпретации, что позволяет быстрее получить первые результаты классификации и понять специфику задачи. Также в качестве потенциальных классификаторов для системы распознавания эмоций по речевому сигналу рассматриваются нейросетевые модели, поскольку они обладают такими важными преимуществами, как устойчивость к шумам во входных данных, довольно высокая степень классификация, достижимая за счёт выбора архитектуры нейронной сети, подходящей для решения данной конкретной задачи.

В третьей главе рассматриваются методы для обработки цифровых сигналов, используемые на этапах предварительной обработки и вычисления характерных особенностей речи в системах распознавания эмоций по речевому сигналу. Приводится обоснование выбора быстрого преобразования Фурье и окна Хэмминга в качестве основных методов для применения на этапе предварительной обработки речевого сигнала. Выполняется обзор методов, используемых для вычисления характерных особенностей речевых сигналов, таких как частота основного тона, форманты, мел-кепстральные коэффициенты, энергия сигнала,

энергия определённых полос частот, скорость пересечения нуля. Делается вывод о том, что основное влияние на точность распознавания эмоций классификатором оказывают мел-кепстральные коэффициенты и форманты.

В четвёртой главе формулируются функциональные требования к разрабатываемому программному средству на основании анализа программ-прототипов из главы 1, приводятся результаты проектирования архитектуры разрабатываемого программного средства. В качестве базы данных для исследований выбирается RAVDESS, даётся характеристика формату файлов, выбранной базы данных. Разрабатывается наиболее подходящая архитектура многослойного персептрона для решения задачи распознавания эмоций по речи. Рассматривается алгоритм вычисления мел-кепстральных коэффициентов в связи с тем, что MFCC коэффициенты оказывают наибольшее влияние из описанных значимых характеристик речевого сигнала на точность работы классификатора. При выборе языка программирования предпочтение отдаётся Python-у, так как данный ЯП является высокоуровневым языком программирования общего назначения, ориентированным на повышение производительности разработчика и читаемости кода, к тому же для языка Python написано колоссальное количество высокопроизводительных библиотек и фреймворков широкой направленности с открытым исходным кодом.

В пятой главе производится поиск оптимальной конфигурации спроектированной системы распознавания эмоций по речевому сигналу на основании выбора различных подмножеств значимых признаков речевого сигнала для расчёта и модификации реализованных моделей классификации. Результатом этого является выбор метода вычисления мел-кепстральных коэффициентов для использования на этапе выделения характерных особенностей речи и пятислойного многослойного персептрона в качестве классификатора. Разрабатываются методы аугментации речевых данных для увеличения размера тестовой базы данных с целью повышения точности работы классификаторов. Даётся оценка эффективности созданной системы распознавания эмоций по речевому сигналу в оптимальной конфигурации в сравнении с уже существующими системами. На основании чего делается вывод о том, что спроектированная система распознавания эмоций по речевому сигналу работает на уровне большинства подобных систем, а некоторые даже превосходит.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Произведён обзор методов, используемых для вычисления характерных особенностей речевых сигналов: частоты основного тона, энергии сигнала, джиттера и шиммера, скорости пересечения нуля, мел-кепстральных коэффициентов. Произведена оценка целесообразности вычисления всей совокупности характерных особенностей речевых сигналов.

2. Исследованы аспекты и целесообразность применения алгоритмов машинного обучения, методология внедрения CrispDM, категории практических задач, решаемых с помощью машинного обучения. Произведён сравнительный анализ эффективности алгоритмов машинного обучения на примере метода опорных векторов, деревьев решений, многослойного персептрона и вероятностных нейронных сетей для решения задач распознавания эмоций.

3. Разработаны методы аугментации речевых данных для увеличения размера тестовой выборки в случае небольшого размера исходных тестовых данных.

4. Спроектирована система распознавания эмоций по голосу на основе наиболее эффективных методов выделения значимых признаков речи и моделей классификации, выявленных в ходе сравнительного анализа.

5. Разработана архитектура программного средства для распознавания эмоций по речевому сигналу. Предложен и обоснован выбор технологического стека для реализации поставленных задач.

6. Программное средство в перспективе может быть усовершенствовано. Под усовершенствованиями в данном контексте можно понимать расширение функциональности, расширение списка рассчитываемых характерных особенностей речевых сигналов, добавление новых классификаторов, сбор и использование новых статистических данных для обучения классифицирующей модели.

Рекомендации по практическому использованию результатов

1. Результаты в виде разработанного ПС могут быть использованы для решения задач определения эмоциональной окраски речевых сигналов.

2. Результаты в виде описания и программной реализации алгоритмов обработки и расчёта характерных особенностей речевых сигналов могут быть использованы в других программных средствах, нуждающихся в алгоритмах для работы с аудиофайлами.

3. Результаты в виде описания и программной реализации моделей классификации эмоций по характерным особенностям речевых сигналов могут использоваться в других работах и программных средствах, исследующих и определяющих эмоциональную окраску речевых сигналов.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Брановицкий, А. А. Распознавание эмоций по голосу на основе машинного обучения / А. А. Брановицкий // Компьютерные системы и сети: материалы 54-й научной конференции аспирантов, магистрантов и студентов БГУИР, Минск, 23-27 апр. 2018 – Минск, 2018. – С. 52.

2-А. Брановицкий А. А. Система автоматического распознавания эмоций по речевому сигналу / А.А. Брановицкий, Д.В. Борисов // Системы обработки информации: материалы международной научной конференции «Информационные технологии и системы» БГУИР, Минск, 30 окт. 2019 – Минск, 2019. – С. 96-97.

Библиотека БГУИР