МИНИСТЕРСТВО НАУКИИ ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования «САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра медицинской физики

Разработка и анализ алгоритма с использованием искусственных нейронных сетей для диагностики психоэмоционального состояния человека по голосовым характеристикам

АВТОРЕФЕРАТ

ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ БАКАЛАВРА

студентки 4 курса 4021 группы

направления 03.03.02 Физика, Института физики Лебединской Анны Андреевны

Научный руководитель доцент, к.фм.н.,	1/0 906.2022	А.Э. Постельга
должность, уч. ст., уч. зв.	личная подпись, дата	инициалы, фамилия
Зав. кафедрой д.фм.н., профессор	AGN 9.06.22	А.В. Скрипаль
должность, уч. ст., уч. зв.	личная подпись, дата	инициалы, фамилия

Введение. Распознавание эмоционального состояния человека в настоящее время представляет немалый интерес для исследований во многих странах. Действующие системы исследования эмоционального состояния человека по голосу различаются видами распознаваемых эмоций, типами используемых баз данных, акустическими параметрами и их производными, а также алгоритмами классификаторов. Эти различия делают результаты распознавания впрямую несопоставимыми.

Системы, распознающие и анализирующие эмоциональное состояние человека по его голосу, могут быть применены в интерактивном телевидении, криминалистике, виртуальном обучении, в работе с детьми, при исследовании нарушений функций мозга, а также будут полезны людям, имеющим какие-либо речевые отклонения.

Цели и задачи работы:

Цель работы: разработать алгоритм с использованием искусственных нейронных сетей для диагностики психоэмоционального состояния человека по физическим характеристикам его голоса.

Задачи работы:

- 1. Исследовать специфические характеристики речи человека, включая особенности женских и мужских голосов.
- 2. Провести критический анализ известных методов диагностики психоэмоционального состояния человека и выявить их достоинства и недостатки.
- 3. Провести экспериментальные исследования фрагментов речи женщин и мужчин.
- 4. Разработать способ разделения преобразованных речевых фрагментов. Основное содержание работы. Основная часть состоит из двух глав. В первой главе рассматриваются современное состояние исследований в области оценки психоэмопионального состояния человека.

В разделе один первой главы приведены основные понятия по данной теме. Во втором разделе первой главы рассматривается взаимосвязь речи и

эмоций человека.

Эмоции – это реакции человека и животного на воздействие внутренних и внешних раздражителей, которые имеют ярко выраженную субъективную окраску и охватывают все виды чувственности и переживаний.

Различают две основные группы эмоций: положительные и отрицательные.

Положительные эмоции — эмоции, связанные с удовлетворением тех или иных потребностей организма. Они вызываются возбуждением центров удовольствия центральной нервной системы.

Отрицательные эмоции – эмоции, связанные с неудовлетворением тех или иных потребностей организма. Отрицательные эмоции вызываются возбуждением центров неудовольствия центральной нервной системы.

При решении вопроса классификации эмоций, необходимо отделять сами эмоции от других эмоционально окрашенных состояний, настроений, аффектов, эмоциональных черт.

Эмоции позволяют правильно трактовать, как говорящий оценивает ситуацию, и какие ответные реакции от него можно ожидать. Если уметь правильно интерпретировать эмоции, то можно снижать их интенсивность, направлять или превращать одну эмоцию в другую. Доказано, что и мысли, и эмоции оказывают влияние на дыхание, выражение лица, положение тела, тон и темп голоса человека. Голос отражает все сильные эмоции человека, он приобретает музыкальный мелодичный характер, меняясь по громкости и по качеству, то есть – по резонансу, тембру, высоте звука. Темп речи и её акцентное членение паузами и логическим ударением по аналогии помогают уловить и Быстрота правильно **ТКНОП** ЭМОЦИИ человека. речи обусловлена индивидуальными качествами и намерениями говорящего, а тембр зависит от того, в каком настроении человек и какое чувство он испытывает по отношению к тому, о чём говорит.

Считается, что спектр эмоций женщин шире. Дело в том, что женщины примерно одинаково используют левое и правое полушария мозга, в то время как

мужчина больше использует именно левое полушарие. Левое полушарие мозга отвечает за логическое мышление и навыки решения проблем, в то время как правое полушарие отвечает за просодические языковые функции, креативность, восприятие лиц и генерирование эмоций. Поскольку мужчины чаще используют левое полушарие мозга, они более эффективно справляются с проблемами, применяя рациональный и логический подход. Женщины же используют оба полушария мозга, поэтому действуют с креативностью, при этом обращаясь к эмоциям. Также правое полушарие мозга отвечает за способность человека выражать свои чувства и мысли, а также за понимание языков. Кроме того, у женщин гораздо более развитая лимбическая система. Она отвечает за такие аспекты жизни человека, как поведение, эмоции и память и позволяет человеку лучше выражать свои чувства и лучше формировать связи с другими людьми. Однако проблема проявляется в том, что она открывает двери для депрессии, особенно по время гормональных скачков, которые появляются во время менструального цикла или беременности. Таким образом, в эти периоды женщина может казаться слишком эмоциональной или подавленной.

В третьем разделе первой главы рассмотрены современные устройства и методы распознавания психоэмоционального состояния человека.

Во второй главе описан результат разработки алгоритма, автоматически классифицирующего речевые фрагменты.

В первом разделе второй главы проведен анализ фрагментов речи женщин и мужчин.

Для определения соотношения психоэмоционального состояния человека и его голосовых физических характеристик, был проведен эксперимент с записью голосов людей и определению истинности их ответов.

Состояние психоэмоционального напряжения, переживаемое человеком, который говорит неправду, отражается на тех характеристиках речи, которые он в наименьшей степени контролирует. Такими физическими характеристиками могут быть интонация, темп речи, тембр голоса, изменение скорости ответов. У

человека, маскирующего или искажающего истинный ответ может появиться дрожь в голосе, нетипичная интонационная окраска речи и долгие паузы.

Для качественного тестирования был проведён опрос, включающий в себя десять вопросов, которые не повторялись и были озвучены человеку впервые после предварительного анкетирования. Вопросы составлены таким образом, что часть ответов оказываются обязательно ложными, поскольку различающиеся фрагменты сформулированы индивидуально:

Каждому из опрошенных была поставлена задача, отвечать словом «правда» на все заданные вопросы. Все тестирования были записаны на диктофон для последующих исследований голосовых сигналов отвечавших.

Для экспериментального исследования записей в среде Python были простроены зависимости громкости голосов отвечавших людей от времени. По ним можно вручную определить такие физические параметры голоса как длительность, громкость ответа, силу голоса и различить физические особенности речи: интонационное ударение и темп речи.

Для разделения речи тестируемых был написан код программы в среде Python. Для реализации алгоритма было решено воспользоваться машинным обучением. Его преимущество в том, что машинное обучение может решать задачи, не проникая в суть предмета.

При помощи самообучающейся искусственной нейросети научились по результатам компьютерной системы распознавать восемь эмоциональных оттенков голоса — нейтрального, спокойного, счастливого, грустного, злого, испуганного, а также демонстрирующего отвращение или удивление.

Применяемая для этой цели нейросеть представляет собой сверточную сеть глубокого обучения с архитектурой VGG-16. Для упрощения работы с голосом, используют спектрограммы сигналов.

Компьютер правильно определяет эмоцию с точностью в среднем 70%, однако для всех эмоции этот показатель точности различный. Лучше всего

автоматизированная система предсказывает нейтральное и спокойное эмоциональное состояние говорящего. А вот счастливую интонацию компьютер во многих случаях путает с испуганной или печальной, так же как удивление часто может распознать как отвращение.

Хоть понимание устной речи и изложение ее в письменном виде уже давно не представляет трудности для компьютера, но имеет значение то, какой смысловой нагрузкой может наделяться речь. Содержание, контекст одного и того же слова или целого высказывания может варьироваться в зависимости от эмоций в голосе говорящего, и эти оттенки смысла компьютер не способен обнаруживать. На вопрос «Всё ли в порядке?» человек отвечает «Конечно». Но как мы можем предположить, смысл слова «конечно» меняется в зависимости от того, каким тоном было произнесено слово — спокойно, вызывающе, жизнерадостно и т. д. Таким образом, употребляя, казалось бы, одно и то же слово, человек может дать совершенно разные ответы на конкретный вопрос.

С целью упрощения звукового анализа, преобразовывают речь в изображения, так называемые спектрограммы. Спектрограммы — это графическое представление зависимости спектральной плотности мощности сигнала от времени.

Достоинством данного метода является достаточно точный результат угадывания эмоций - 70%. однако ограниченный и неизменный набор распознаваемых эмоций (всего восемь) – это недостаток программы.

В разделе номер два и номер три второй главы изучены понятия машинное обучение и свёрточные нейронные сети.

Машинное обучение — это создание и применение моделей, извлеченных из данных в результате обучения. Где в качестве модели рассматриваются подробное описание математической или вероятностной связи, существующей между различными величинами.

Существуют следующие типы обучения:

- Обучение с учителем
- Обучение без учителя

• Обучение с подкреплением.

Обучение без учителя предполагает использование данных для моделирования, поступающих без соответствующих категорий и обработки. При достаточном объеме данных можно находить паттерны и структуры. Два наиболее мощных инструмента, которые используют специалисты для машинного обучения на основе только исходных данных, это кластеризация и уменьшение размерности.

Кластеризацией называют процесс разделения данных на отдельные сегменты сходных элементов. В этом отношении кластеризация похожа на классификацию данных без каких-либо соответствующих меток классов.

Алгоритмы машинного обучения дают хороший результат при решении определённых задач. Однако, попытки пошагово, с применением дедуктивных методов, проследить, как получился такой результат, нерезультативны. Тщательно продуманная система машинного обучения изучает тысячи параметров, но выяснение значения каждого параметра не всегда является её основной задачей.

Для реализации алгоритма, автоматически классифицирующего речевые фрагменты наиболее подходящий тип обучения без учителя. А для удобства исследования речевых фрагментов, их лучше всего представить в виде изображений. С задачей распознавания и классификации изображений хорошо справляются свёрточные нейронные сети.

Основная идея, лежащая в основе сверточных нейронных сетей, состоит в том, что вполне достаточно локального осмысления изображения. Практическое преимущество сверточных нейронных сетей таково, что, имея несколько параметров, можно значительно сократить время на обучение, а также объем данных, необходимых для обучения модели.

Стоит отметить, что если входными данными являются звуковые данные или изображения, то сверточные нейронные сети оказываются справляются весьма успешно. Для промышленности особый интерес представляют изображения. Например, при регистрации в соцсети вы загружаете фотографию

профиля, а не запись своего приветствия.

Сверточная нейронная сеть состоит из нескольких слоев сверток и подвыборок. Слой свертки дает различные представления изображения, тогда как подвыборка слоя упрощает расчеты благодаря снижению размерности без значительной потери информации. Изображение, пропускается через серию свёрточных, нелинейных и других слоев и генерируется вывод. Выводом может быть класс или вероятность классов, которые лучше всего описывают изображение.

Машинное обучение представляет собой вечную борьбу за разработку модели, которая обладала бы достаточной выразительностью для представления данных, но при этом не была такой универсальной, чтобы доходить до переобученности и запоминать паттерны. Нейронные сети предлагаются как способ повышения выразительности; хотя, как можно догадаться, они сильно страдают от ловушек переобучения. Переобучение возникает, когда обученная модель исключительно точна на обучающем наборе данных и плоха на проверочном наборе данных. Эта модель, вероятно, чрезмерно универсальна для небольшого объема доступных данных, и в конце концов она просто запоминает обучающие данные.

Поэтому, чтобы увидеть, насколько правильно работает свёрточная нейросеть, необходимо провести тестирование точности после завершения обучения на другом наборе данных.

В четвёртом разделе второй главы рассмотрен принцип преобразования аудиофайлов в изображения. Для решения задачи преобразования аудиозаписей была написана программа в среде Python.

В состоянии напряжения, когда человек говорит неправду скорость и диапазон изменений частоты основного тона существенно возрастают. А между собой не согласовываются аналогичные изменения характерные для интенсивности и громкости речи.

В пятом разделе второй главы описаны устройство и работа нейросети.

Для создания и запуска работы нейросети необходимо выполнить

следующие операции:

- Построение модели
- Настройка слоев
- Компиляция модели
- Обучение модели
- Оценка точности работы
- Создание предсказания

Для построения нейронной сети необходимо настроить слои модели, а затем скомпилировать модель. Основным строительным блоком нейронной сети является слой. Слои извлекают представления из введенных в них данных.

Но, прежде чем модель будет готова к обучению, ей нужно еще несколько настроек. Они добавляются на этапе компиляции модели.

Функция потерь — измеряет, насколько точна модель во время обучения. Необходимо минимизировать эту функцию, чтобы "направить" модель в правильном направлении. Оптимизатор — с его помощью модель обновляется на основе данных и функции потерь. Метрики — используются для мониторинга этапов обучения и тестирования.

При разработке модели как правило делают разделение проверки: используют 80% изображений для обучения и 20% - для проверки.

Для обучения модели нейронной сети необходимо выполнить следующие шаги. Подать обучающие данные к модели. Модель учится связывать изображения и метки. Далее запрос модели сделать прогноз относительно тестового набора для того, чтобы убедиться в соответствии прогнозов меткам.

Затем нужно сравнить, как модель работает с тестовым набором данных. Обычно точность на тестовом наборе данных немного меньше, чем точность на обучающем наборе данных. Этот разрыв между точностью обучения и точностью теста представляет собой переоснащение. Переоснащение происходит, когда модель машинного обучения работает хуже с новыми, ранее невиданными входными данными, чем с обучающими данными. Переобученная модель «запоминает» шум и детали в обучающем наборе данных до такой

степени, что это негативно влияет на производительность модели на новых данных. Обученную модель можно использовать для прогнозирования некоторых изображений.

Заключение. В ходе работы были исследованы специфические характеристики речи человека и проведен критический анализ известных методов диагностики психоэмоционального состояния человека. Проведены экспериментальные исследования фрагментов речи женщин и мужчин и разработан способ разделения преобразованных речевых фрагментов по заданным категориям с использованием машинного обучения.

В данной работе задача распознавания ложного высказывания по голосу человека, который находится в состоянии психоэмоционального напряжения, реализована с помощью алгоритма из трёх этапов. Создание аудиозаписей с ответами тестируемых людей и преобразование голосовых сигналов в изображения. Затем классификация этих изображений на две категории «Ложь» и «Истина» с использованием свёрточной самообучающейся нейронной сети.