МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**Национальный исследовательский ядерный университет «МИФИ»**

**Институт   
интеллектуальных кибернетических систем**

**Кафедра кибернетики (№ 22)**

Направление подготовки 09.04.04 Программная инженерия

**Расширенное содержание пояснительной записки**



к научно-исследовательской работе студента на тему:

Программная реализация алгоритмического обеспечения для решения задачи отображения пространств эмоций различных размерностей.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Группа | М21-534 | | |  | | |
| Студент |  |  |  | |  | Чудновец И.В. |
|  |  |  | (подпись) | | (ФИО) | |
| Руководитель |  |  |  | |  | Климов В.В, |
|  | (0-20 баллов) |  | (подпись) | | (ФИО) | |
| Научный консультант |  |  |  | |  | Тихомирова Д.В. |
|  | (0-20 баллов) |  | (подпись) | | (ФИО) | |

**Москва 2022**

# **Реферат**

Пояснительная записка содержит n страниц, m рисунков, l таблиц.

Количество использованных источников k.

Ключевые слова: пространство эмоций, базовые эмоции, PAD, FACS, action units, нейронные сети, регрессионные модели, web-сервис, виртуальные ассистенты.

Целью данной работы является продолжение разработки модуля, предназначенного для отображения между пространствами эмоций различных размерностей, в виде реализации web-сервиса для удобной демонстрации работы обученных моделей из предыдущего семестра.

В первом разделе описываются виртуальные ассистенты в целом, рассматриваются вопросы представления эмоций в виде формальных моделей, описана общая постановка задачи работы.

Во втором разделе выбирается стек технологий для реализации приложения, рассматриваются паттерны проектирования, полезные для данного ПО, а также разрабатываются форматы файлов обученных моделей (1 семестра и будущих семестров).

В третьем разделе представлены вопросы проектирования web-приложения и его клиентской части (web-сервиса), служащих удобным программным интерфейсом для работы с обученными моделями.

В четвёртом разделе изложена программная реализация web-сервиса и демонстрация его основных возможностей.

В заключении подводятся итоги проведённой работы с кратким описанием результатов по каждому разделу.

**СОДЕРЖАНИЕ**

[**Реферат**](#_30j0zll) **2**

[**Введение**](#_3znysh7) **4**

[**Раздел 1. Аналитическая часть.**](#_tyjcwt) **6**

[1.1 Технологии и подходы для реализации виртуальных ассистентов.](#_3dy6vkm) 6

[1.2 Изучение и анализ литературы на тему представления эмоций в виде формальных моделей.](#_1t3h5sf) 8

[1.2.1 Общие характеристики дискретных и многомерных моделей.](#_4d34og8) 8

[1.2.2 Модель эмоционального состояния PAD.](#_2s8eyo1) 11

[1.2.3 Система кодирования лицевых движений FACS.](#_17dp8vu) 12

[1.3 Постановка задачи, требования к разрабатываемому ПО.](#_3rdcrjn) 14

[**Раздел 2. Теоретическая часть**](#_35nkun2) **16**

[2.1 Проведение сравнительного анализа регрессионных и нейросетевых методов.](#_1ksv4uv) 16

[2.2 Выбор датасета, анализ данных, формирование обучающей и тестовой выборок.](#_44sinio) 17

[2.2.1 Выбор датасета.](#_2jxsxqh) 17

[2.2.2 Анализ данных.](#_z337ya) 17

[2.2.3 Формирование обучающей и тестовой выборок.](#_3j2qqm3) 21

[2.3 Разработка алгоритма обучения моделей.](#_1y810tw) 22

[**Раздел 3. Инженерная часть**](#_2xcytpi) **23**

[3.1 Проектирование архитектуры приложения. Описание архитектуры приложения.](#_1ci93xb) 23

[3.2 Проектирование макетов клиентской части приложения.](#_3whwml4) 23

[3.3 Проектирование серверной части web-приложения.](#_2bn6wsx) 23

[**Раздел 4. Технологическая и практическая часть**](#_1pxezwc) **24**

[4.1 Обучение моделей на обучающей и тестирование на тестовой выборках.](#_49x2ik5) 24

[4.2 Оценка точности работы моделей.](#_2p2csry) 24

[4.3 Выбор стека технологий для реализации приложения.](#_147n2zr) 24

[4.4 Реализация клиентской и серверной частей web-сервиса.](#_3o7alnk) 24

[**Заключение**](#_ihv636) **25**

[**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**](#_1hmsyys) **26**

# 

# **Введение**

Существуют различные подходы к описанию эмоций человека. Одни модели включают в себя фиксированное количество базисных (фундаментальных) эмоций, другие используют ограниченное число переменных (осей), которые задают эмоциональную точку в пространстве. Первые модели называются дискретными, а вторые – многомерными [1].

Исследователи, разделяющие дискретную теорию эмоций, расходятся во мнениях относительно количества и содержания фундаментальных эмоций. Сторонники этого подхода сосредоточены на изучении внутренней организации индивидуальных эмоций, а не на их взаимосвязи. Наиболее известным психологом в этой области считается Пол Экман, выделяющий 6 базовых эмоций — гнев, удивление, страх, радость, грусть и отвращение [1]. В рамках данной работы мы будем считать, что базисных эмоций всего 7 (также присутствует нейтральная эмоция).

Многомерные модели эмоциональных состояний различаются между собой количеством измерений (осей) и их смысловой нагрузкой. В данном исследовании мы будем рассматривать многомерную модель PAD [2], в которой всего 3 измерения: P – Pleasure (удовольствие), A – Arousal (возбуждение), D – Dominance (доминирование).

Полом Экманом в XX веке была разработана система кодирования лицевых движений (FACS) [3]. Эта система позволяет классифицировать выражения лица человека на основе его мимики. С использованием FACS можно закодировать практически любое, анатомически возможное выражение лица, деконструируя его на вызвавшие его конкретные двигательные единицы (action units) и их временные сегменты [4].

С помощью двигательных единиц из FACS Полом Экманом и Уоллесом Фризеном была создана эмоциональная система кодирования лицевых движений (EmFACS), которая рассматривает только лицевые движения, связанные с эмоциями. Стоит отметить, что в научном мире нет единого мнения на тему однозначного соответствия между мимикой и эмоциями для всех людей [5]. Таким образом нельзя с полной уверенностью сказать, что, например, выражение лица любого человека, который испытывает страх выглядит именно так, а никак иначе. Однако, большинство людей в современном обществе выражает эмоции приблизительно одинаково, что даёт возможность использовать труды Пола Экмана в данной работе.

С учётом вышесказанного, встаёт вопрос о необходимости иметь возможность отображать пространства эмоций различных размерностей друг в друга.

# **Раздел 1. Аналитическая часть.**

## 1.1 Технологии и подходы для реализации виртуальных ассистентов.

Виртуальный ассистент [6] (англ. Virtual assistant) — программный агент, который может выполнять задачи (или сервисы) для пользователя на основе информации, введенной пользователем, данных о его местонахождении, а также информации, полученной из различных интернет-ресурсов (погода, уличное движение, новости, курсы валют и ценных бумаг, розничные цены в магазинах и т. д.). Примерами таких агентов являются программы Siri, Google Assistant (Google Now), Amazon Alexa, Microsoft Cortana, Bixby, Voice Mate, Алиса, Маруся, Дуся, Салют и другие.

Искусственный интеллект уже стал частью нашей повседневной жизни. В каждом мобильном телефоне есть умный голосовой помощник, который расскажет о погоде, подскажет, где ближайший банкомат, вызовет такси или закажет билеты в кино. Но это только один из примеров виртуального помощника, самый известный и знакомый. На самом деле у таких решений гораздо больше областей применения.

Так что же такое виртуальный помощник (ассистент)? В общем смысле это сервис (программа, мобильное приложение) — выполняющий действия в ответ на информацию, введенную пользователем и/или найденную в Сети. Кроме того, искусственный интеллект избавляет человека от необходимости самостоятельно выполнять некоторые операции – как правило, простые, то есть такие, которые под силу роботу.

Возможности виртуальных помощников используются во всех сферах жизни. В автомобили внедряются голосовые помощники, чтобы владелец мог без помощи рук открывать и закрывать двери, запускать двигатель, узнавать уровень топлива в бензобаке. Роботы управляют устройствами умного дома, их устанавливают в гостиницах, например, чтобы гости могли заказывать услуги в номер с помощью простых команд.

Основная функция голосовых виртуальных ассистентов – имитировать реальное общение. Уровень развития этой функции у цифровых помощников разный. Некоторые роботы способны поддерживать только самые простые диалоги, читая предварительно записанные звуковые сообщения. Другие неотличимы от человека: понимают смысл обращенной к ним речи, знают синонимы, отвечают без промедления, замолкают, когда их перебивают, и даже употребляют слова-паразиты. В результате собеседник даже не всегда может определить, что ведет диалог с роботом. Понятно, что такие возможности могут быть очень полезны для бизнеса.

Кроме коммуникационных функций, виртуальные ассистенты (голосовые боты) могут выполнять различные действия по заданным сценариям: планировать маршрутизацию звонков, отправлять SMS, записывать результаты вызовов в виде отчетов, использовать рабочие данные по информационным ресурсам компании и даже фиксировать изменения в них – при условии, что робот интегрирован с нужными системами.

Интернет-магазины с помощью голосовых ботов сообщают об акциях и специальных предложениях, регистрируют клиентов в программе лояльности, активируют корзину, увеличивают продажи, помогают клиентам с заказом.

Виртуальный помощник в салоне красоты или другой организации из сферы услуг напомнит клиенту о приеме, при необходимости перенесет время визита, в результате чего показатель «рентабельности» повысится.

Службы доставки посредством цифровых помощников информируют получателей о статусе заказа, согласовывают доставку.

Банки, страховые компании, операторы связи и интернет-провайдеры используют виртуальных помощников, чтобы оповещать о новых услугах и тарифах, напоминать клиентам о задолженности и облегчить совершение платежей.

И это не все области, где используются виртуальные помощники, и не все решаемые ими задачи. Практически в любой компании виртуальный ассистент может выполнять функции технической поддержки: отвечать на частые вопросы клиентов, консультировать по несложным задачам, перенаправлять звонок оператору. Кроме того, его можно использовать для проведения опросов и анкетирования, например, чтобы узнать мнение клиентов о новом продукте. Виртуальные помощники также помогают во внутрикорпоративных коммуникациях: решают текущие вопросы сотрудников, участвуют во взаимодействии отделов.

Ассистент может быть оптимизирован для своего рабочего канала. Как правило, виртуальные помощники используются в голосовых каналах (на телефоне) и в интернет-каналах (на веб-сайтах). Первый тип относится в основном к голосовым помощникам, ко второму – чат-ботам. У них разные сильные и слабые стороны, которые необходимо учитывать в реальных проектах. Хороший результат для голосового канала может не подойти для интернет-чата и наоборот.

Указанные выше ассистенты (голосовые помощники и чат-боты) смогли быть реализованы в связи с решением главных задач, лежащих в основе этих ассистентов. Например, внутри голосового помощника заложены алгоритмы распознавания и синтеза речи. Впервые эти задачи были реализованы в коммерческих программах распознавания речи в 90-х годах [7][8]. Чат-боты же основаны на заранее определенных бизнес-правилах, определяющих ход диалога между пользователем и ботом. Как правило, чат-боты избегают вопросов, требующих свободного ответа, а ставят их в виде «тестового» варианта ответа (с конечным количеством вариантов). Эта технология построения интеллектуаьных систем хорошо изучена и уже нашла широкое применение на практике [9].

Также, перспективной ветвью развития виртуальных ассистентов могут стать помощники, использующие цифровых аватаров [10] – виртуальных персонажей, имеющих внешний облик, схожий с человеком. Данный вид ассистентов находится сейчас на ранних этапах своего развития, т.к. часть задач, лежащих в их основе не имеет достаточной фундаментальной базы. Так, задача распознавания лиц уже имеет коммерческие решения на данный момент [11], но она является лишь частью необходимой базы для создания виртуальных ассистентов с цифровыми аватарами. Продолжаются работы по синтезу выражений лиц для виртуальных персонажей, т.к. в качественном ассистенте, мимика и эмоции аватаров должны быть близки к реальным для людей, что на данный момент далеко от желаемого результата.

## 1.2 Изучение и анализ литературы на тему представления эмоций в виде формальных моделей.

Много лет учёные всего мира работали над задачей распознавания мимики человека. В наши дни уже есть успехи в этом направлении. Если раньше для разложения выражения лица на составляющие в FACS требовался специально обученный человек, то теперь нейросети могут делать это автоматически.

### 1.2.1 Общие характеристики дискретных и многомерных моделей.

Есть два принципиально разных подхода к изучению организации эмоций. С точки зрения дискретной модели [1] эмоциональная сфера состоит из определенного количества первичных, базовых или фундаментальных эмоций. Комбинации первичных эмоций - это эмоции второго порядка. Статус первичных эмоций обычно приписывается гневу, страху, грусти, счастью и т.д. Сторонники этого подхода сосредоточены на изучении внутренней организации индивидуальных эмоций, не уделяется внимание взаимосвязям между эмоциями.

В многомерной модели акцент делается на многомерном пространстве эмоций, образованном ограниченным числом шкал (осей), например, негативная - позитивная, сильная - слабая, активная - пассивная. Такие модели ориентированы на изучение сходства и различий между эмоциями.

Исследователи, разделяющие дискретную теорию эмоций, расходятся во мнениях относительно количества и принципов фундаментальных эмоций. Так, П. Экман и его сотрудники на основе изучения выражения лица выделяют 6 основных эмоций: гнев, страх, отвращение, удивление, печаль и радость. Р. Плутчик рассматривает 8 основных эмоций: одобрение, гнев, ожидание, отвращение, радость, страх, печаль, удивление. Классификация К. Изарда с 10-ю основными эмоциями, стала очень популярной. Он отмечает гнев, презрение, отвращение, дистресс (горе - страдание), страх, вину, интерес, радость, стыд, удивление. Есть и более простые классификации. Например, Я. Панксеп различает всего 5 основных эмоций: страх - тревога, ярость - гнев, паника - горе, ожидание - исследование, радость - игра.

Дж. Грей исследует три основные эмоции: ярость - ужас, тревогу и радость. По мнению Р. Мауэра, вообще следует говорить только о двух первичных эмоциях - удовольствии и боли.

Для дискретной модели вопрос о составе фундаментальных эмоций является критическим. Вопрос в том, сколько всего эмоций. Ответ зависит от того, какие показатели, какие реакции использовались при исследовании эмоций (вегетативные, биохимические, ЭМГ мышц лица и т.д.). По выражению лица сложно выделить более 10 паттернов активации мимических мышц, а по вегетативным параметрам можно описать еще меньше комплексов, связанных с эмоциями.

Многомерные модели эмоциональных состояний различаются между собой количеством измерений (осей) и их смысловой нагрузкой.

Адекватным методами для многомерной модели являются методы многомерного шкалирования (ММШ) или факторный анализ.

ММШ впервые был протестирован для изучения распознавания эмоций другого человека по внешним экспрессивным реакциям лица. Первоначальная процедура масштабирования применяет оценку внешних проявлений эмоций только по внутренним, заранее определенным шкалам. С развитием методов многомерной статистики в качестве исходных данных стали использоваться субъективные оценки различия или сходства лиц, выражающих различные эмоциональные переживания. Это позволяет определить и интерпретировать минимально допустимое количество шкал (факторов), которые ориентированы на предмет при вынесении суждений о различии, и тем самым приблизиться к механизму декодирования эмоций. В многомерном пространственном представлении каждое выражение представляет собой точку со своими собственными координатами или вектором, компоненты которого соответствуют вкладам главных осей (масштабов) пространства.

К первым работам в этом направлении относятся исследования Р. Вудворта (Woodworth R.S.) и Г. Шлосберга (Schlosberg N.). Р. Вудворте предложил первую классификацию выражений эмоций на лице с использованием шестичленной линейной шкалы: любовь, счастье, веселье; удивление; страх, страдание; гнев, решимость; отвращение; презрение.

Г. Шлоссберг, применив классификацию Вудворта к анализу фотографий людей с разной мимикой, обратил внимание на частое смешение самых далеких групп внешних выражений эмоций. Позже, в рамках теории активации Д. Линдсли (D. Lindsley), Г. Шлоссберг добавил третье измерение: сон - напряжение, создаваемое, таким образом, тремя измерениями эмоций, предложенными В. Вундтом в 1896 году: удовольствие - неудовольствие, расслабление - напряжение, спокойствие – волнение.

Однако в работе П. Абельсона и В. Сермата (Abelson R., Sermat V.), а также Р. Шепарда (Chapard R.) с использованием ММШ, когда испытуемые оценивали степень схожести между всеми возможными комбинациями пар из 13 фотографий по 9-балльной системе, были выдвинуты аргументы в пользу двухмерного решения. Была выделена шкала знаков эмоций (удовольствие - неудовольствие) и шкала активации (сон - бодрствование). Эти два фактора составляют примерно 75% от общего веса всех факторов. Аналогичные результаты были получены с использованием других наборов фотографий. В более поздних работах некоторые исследователи выделили новые измерения, такие как контроль - импульсивность, внимание - невнимательность, уверенность - неуверенность, которые были предложены в качестве третьей шкалы или вместо шкалы активации.

В экспериментах Дж. Рассела и М. Буллока (Russal J., Billok M.) с ранжированием мимики по фотографиям принимали участие дети 2, 3, 4 и 5 лет, а также студенты. Для испытуемых каждой возрастной группы было получено двумерное пространство, образованное двумя ортогональными осями: знаками и уровнем активации. Эмоции, изображенные на фотографиях, представлены точками, образующими фигуру, близкую по форме к кругу. Сходство эмоциональных пространств и взрослых детей в возрасте, когда еще нет дифференцированных представлений о различных эмоциях, свидетельствует о том, что структура опосредована восприятием эмоций, а не смысловыми связями, образующимися в процессе усвоения языка. Третья ось со значительно меньшим весом как у детей 4 лет, так и у взрослых была интерпретирована как шкала уверенности - смущения.

Особо следует выделить серию работ Дж. Рассела и его коллег. Применяя ММШ к субъективным оценкам сходства эмоций, он получил двумерную круговую модель эмоционального опыта (A Circumplex model of affect [12]). Эта двухмерная модель является PA-частью модели PAD.

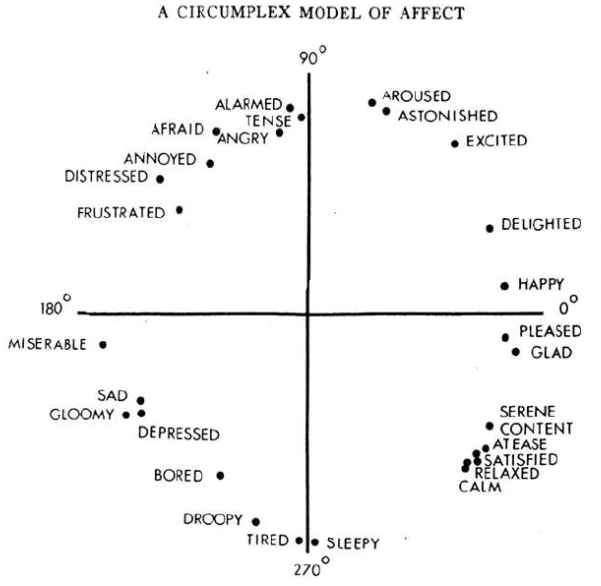


Рисунок 1.1 – A Circumplex model of affect

### 1.2.2 Модель эмоционального состояния PAD.

В данном исследовании мы будем рассматривать многомерную модель PAD [2], в которой всего 3 измерения: P – Pleasure (удовольствие), A – Arousal (возбуждение), D – Dominance (доминирование) для всех представлений эмоций. Модель эмоционального состояния PAD - это психологическая модель, разработанная Альбертом Мехрабианом и Джеймсом А. Расселом (1974 г. и позже) для описания и измерения эмоциональных состояний. Первоначально использовавшаяся в теории психологии окружающей среды, основная идея заключалась в том, что физическая среда влияет на людей через их эмоциональное воздействие. Впечатление было использовано Питером Лэнгом и его коллегами, чтобы предложить физиологическую теорию эмоций. Его также использовал Джеймс А. Рассел для разработки теории эмоциональных эпизодов (относительно коротких эмоционально заряженных событий). PA-часть PAD была встроена в закрытую модель эмоционального опыта, и эти два измерения были названы «базовым аффектом». Часть D PAD была переосмыслена как часть процесса оценки эмоционального эпизода (холодная когнитивная оценка эмоциональной ситуации). Более подробная версия этого подхода называется психологической теорией построения эмоций. Модель PAD (Удовольствие, Возбуждение, Доминирование) использовалась для изучения невербального общения, такого как язык тела в психологии. Он также использует в потребительском маркетинге и создает анимированных персонажей, которые выражают эмоции в виртуальных мирах.

PAD использует трехмерные шкалы, которые теоретически могут иметь любое числовое значение. Размерная структура напоминает работы Вильгельма Вундта 19-го века, который также использовал трехмерную систему, а также работы Чарльза Э. Осгуда 20-го века.

Шкала удовольствия-неудовольствия измеряет, насколько приятно или неприятно человек чувствует что-то. Например, и гнев, и страх - неприятные эмоции, и оба они считаются недовольством. Однако радость - это приятное чувство.

Шкала возбуждения-бездействия измеряет, насколько человек чувствует себя возбужденным или гипнотическим. Дело не в интенсивности эмоций - горе и депрессия могут быть легким возбуждением и сильными чувствами. Хотя и гнев, и ярость являются неприятными эмоциями, ярость имеет более высокую интенсивность или более высокое состояние возбуждения. Однако скука, которая также является неприятным состоянием, имеет низкую ценность для возбуждения.

Шкала «Доминирование-покорность» представляет собой ощущение контроля и доминирования в сравнении с тем, что вы контролируете или подчиняетесь. Например, если страх и гнев - это неприятные эмоции, гнев - это доминирующая эмоция, а страх - это эмоция подчинения.

Более сокращенная версия модели использует всего 4 значения для каждого измерения, обеспечивая только 64 значения для возможных эмоций. Например, гнев - довольно неприятная, довольно возбужденная и умеренно доминирующая эмоция, а скука - слегка неприятная, совсем не возбужденная и в большинстве случаев не доминирующая эмоция.

### 1.2.3 Система кодирования лицевых движений FACS.

Полом Экманом в XX веке была разработана система кодирования лицевых движений (FACS) [3]. Эта система позволяет классифицировать выражения лица человека на основе его мимики. С использованием FACS можно закодировать практически любое, анатомически возможное выражение лица, деконструируя его на вызвавшие его конкретные двигательные единицы (action units) и их временные сегменты [4].

FACS фактически не предусматривает любую биомеханическую информацию о степени активации мышц, но основные мышцы, участвующие в формировании выражений лица, были добавлены сюда для удобства читателя.

Двигательные единицы (action untis) представляют собой основные движения, совершаемые отдельными мышцами или группой мышц.

Двигательные дескрипторы (action descriptors) представляют собой движения, совершаемые группами мышц (например, выдвигание нижней челюсти вперед). Мышечная основа для этих движений не указана.

Для обозначения интенсивности задействования action units, в FACS к номеру AU добавляют латинские буквы с A по E, в зависимости от минимальной — максимальной интенсивности движения. Например, AU 1А является самым трудноразличимым движением AU 1, а AU 1Е является максимально возможной интенсивностью для определенного человека.

* A Слабо различимое
* B Незначительное
* C Заметное или ярко выраженное
* D Сильное или крайне заметное
* E Предельное

Двигательные единицы относительно воображаемой вертикальной оси лица могут быть:

* Двусторонние, симметричные (B — bilateral). Например, (B)1.
* Односторонние (U — unilateral). Например, U12, U14:
* левая (L — left). Например, L12.
* правая (R — right). Например, R14.

Примеры кодирования лицевых движений с помощью FACS можно изучить подробнее по ссылке [3].

С помощью двигательных единиц из FACS Полом Экманом и Уоллесом Фризеном была создана эмоциональная система кодирования лицевых движений (EmFACS), которая рассматривает только лицевые движения, связанные с эмоциями. Стоит отметить, что в научном мире нет единого мнения на тему однозначного соответствия между мимикой и эмоциями для всех людей [5]. Таким образом нельзя с полной уверенностью сказать, что, например, выражение лица любого человека, который испытывает страх выглядит именно так, а никак иначе. Однако, большинство людей в современном обществе выражает эмоции приблизительно одинаково, что даёт возможность использовать труды Пола Экмана в данной работе.

## 1.3 Постановка задачи, требования к разрабатываемому ПО.

С учётом вышесказанного, встаёт вопрос о необходимости иметь возможность отображать пространства эмоций различных размерностей друг в друга. Так, в научно-исследовательской работе 1-го семестра, были обучены нейросетевые модели, решающие задачу отображения из двухмерного пространства в семимерное. По сути, были построены функции для отображения вектора модели эмоционального состояния PA (координатная плоскость pleasure, arousal из пространства PAD) в вектор 7-ми эмоций из непрерывной модели Пола Экмана (Neutral – нейтральная, Happy – радость, Sad – печаль, Angry – гнев, Surprised – удивление, Scared – страх, Disgusted – отвращение).

В дальнейших семестрах планируется обучить модели на решение задач обратного отображения из 7-ми мерного в 2-х мерное. Также будут затронуты вопросы отображения выражений лиц людей закодированных с помощью FACS в многомерную модель PA и обратно, а также из FACS в непрерывную модель Пола Экмана и обратно.

Если в прошлой работе была использована исключительно нейросетевая модель, то далее будут использованы ещё и регрессионные статистические модели. Это разнообразие моделей позволит сравнивать результаты между собой и оценивать их качество.

Целью же данной работы, является упрощение взаимодействия пользователя с обученными моделями. Для достижения этой цели ведётся разработка программного обеспечения (ПО) – web-сервиса, являющегося удобным программным интерфейсом для работы с обученными моделями. Разрабатываемое ПО должно обладать следующими возможностями:

1. загрузка моделей (1 или нескольких)
2. выбор типа преобразования (2 -> 7, 7 -> 2, 42 -> 7, 7 -> 42, 2 -> 42, 42 -> 2)
3. визуализация полученного преобразования входных данных в выходные данные
4. выдача ошибки, при необходимости

Чтобы получить детальное представление о виде графического интерфейса, необходимо разработать макеты клиентской части приложения (пункт 3.2), которые позволяли бы вводить данные в одном эмоциональном пространстве и получать преобразованные данные для другого эмоционального пространства (например, из PA в 7-ми мерное).

Помимо этого, необходимо спроектировать внутреннюю логику web-сервиса (пункт 3.3), разбить приложение на классы, а классы разнести по модулям (пункт 3.1). После этого предстоит запрограммировать данный web-сервис (пункт 4.1).

# **Раздел 2. Теоретическая часть**

## 2.1 Проведение сравнительного анализа регрессионных и нейросетевых методов.

Задача отображения эмоций, поставленная в текущем исследовании в разделе 1.4, напоминает задачу регрессионного анализа, суть которой заключается в установлении формы зависимости между переменными, оценке неизвестных значений (прогноз значений) зависимой переменной. Как следствие, существуют различные методы решения этой задачи как статистические, так и нейросетевые.

Более точно можно классифицировать задачу, как задачу повышения (снижения) размерности. По своему смыслу это преобразование данных, состоящее в уменьшении числа переменных путём получения главных переменных (в случае сжатия). Преобразование может быть разделено на отбор признаков и выделение признаков.

К статистическим методам решения задачи снижения размерности можно отнести метод главных компонент (МГК), линейный дискриминантный анализ (ЛДА) и обобщённый дискриминантный анализ (ОДА), регрессионный анализ. К нейросетевым — применение автокодировщика (автоэнкодера).

В этой работе рассматривается не только применение нейросетевого подхода (как это было в предыдущей работе), но и используются регрессионные модели.

Для нейросетевого расширения (сжатия) данных, кажется подходящей архитектура автоэнкодера. Также могут пригодиться Cамоорганизующиеся сети Кохонена, осуществляющие проецирование многомерного пространства в пространство с более низкой размерностью. Возможно, в решении будет использована Ограниченная машина Больцмана (задача снижения размерности).

Остановимся подробнее на применении автоэнкодеров [15], специального вида нейронных сетей прямого распространения (англ. feed-forward networks) с бутылочным (в виде бутылочного горлышка) скрытым слоем. Обучение глубоких кодировщиков обычно осуществляется с использованием жадного послойного предобучения (например, используя каскад ограниченных машин Больцмана), за которым следует этап тонкой настройки, основанный на методе обратного распространения ошибки.

Автокодировщик может быть использован [15] для изучения функций нелинейного снижения размерности и кодирования вместе с обратной функцией из кодированного к исходному представлению. Основными практическими приложениями автокодировщиков остаются уменьшение шума в данных, а также уменьшение размерности многомерных данных для визуализации. С определёнными оговорками, касающимися размерности и разрежённости данных, автокодировщики могут позволять получать проекции многомерных данных, которые оказываются лучше тех, что даёт метод главных компонент либо какой-либо другой классический метод [16].

Стоит отметить и преимущества регрессионных моделей. Регрессионные модели берут за основу математические формулы и уравнения, в отличии от нейросетевых моделей, которые оперируют архитектурой сети, количеством слоёв и нейронов в слое, а также методом обучения нейросети. Как следствие, подход в регрессионных моделях более наглядный и понятный с точки зрения интерпретации результатов. Этот подход можно охарактеризовать как более научный и фундаментальный. К тому же, в статистике активнее используется априорная информация, которой, как правило, немало, что позволяет большее внимание уделить тщательному анализу задачи и меньшее непосредственному обучению модели.

## 2.2 Выбор датасета, анализ данных, формирование обучающей и тестовой выборок.

В данном разделе выбирается датасет, данные в датасете подвергаются тщательному анализируются, формируются обучающие и тестовые выборки.

### 2.2.1 Выбор датасета.

Всего есть 3 датасета на котором можно обучить нейросеть. Это логи программ FaceReader, OpenFace и iPad (система компьютерного зрения в iOS — ARKit). Выбор пал на логи программы FaceReader, поскольку они содержат всю необходимую информацию для обучения в удобном формате. Все данные разбиты на файлы логов. Каждый лог - отдельный файл с именем участника, временем и датой прохождения им теста. Если собрать данные со всех файлов, то получится > 25 тыс. данных (пар векторов (2-х; 7-ми мерных)).

### 2.2.2 Анализ данных.

Строки логов FaceReader′a содержат 75 столбцов, среди которых стоит выделить:

Video Time - Время с начала видео (тип datetime в формате 'HH:MM:SS.sss')

Neutral - Значение нейтрального эмоционального состояния (тип float ∈ [0; 1]),

Happy - Значение радости (тип float ∈ [0; 1]),

Sad - Значение грусти (тип float ∈ [0; 1]),

Angry - Значение гнева (тип float ∈ [0; 1]),

Surprised - Значение удивления (тип float ∈ [0; 1]),

Scared - Значение страха (тип float ∈ [0; 1]),

Disgusted - Значение отвращения (тип float ∈ [0; 1]),

Valence - Значение валентности (тип float ∈ [-1; 1]),

Arousal - Степень возбуждения (тип float ∈ [0; 1]).

Вышеуказанные данные потребуются нам для текущей работы, преобразования эмоций из 2-мерного пространства векторов (Valence; Arousal) в пространство 7-мерных векторов (Neutral; Happy; Sad; Angry; Surprised; Scared; Disgusted). Время (Video Time) — не играет роли в задачи, но удобно для обработки информации и работы с данными. В следующих семестрах мы будем использовать оставшиеся данные, не представленные в текущем исследовании, (42 столбца Action Units) двигательных единиц из FACS.

Традиционно в модели PAD в координатной плоскости PA (VA в случае с нашим датасетом) точка характеризуется углом наклона (a) луча в полярных координатах (Рисунок 1.1). Так, поделив значение координаты Arousal на Valence можно получить tg(a) угла наклона (Arousal/Valence = tg(a)).

После анализа данных логов FaceReader′а определены примерные расположения семи чистых эмоций (Рисунок 3.1). Так точка центра Neutral эмоции имеет координаты ≈ (0; 0.25) (tg(a) → ∞), Happy ≈ (1; 0.75) (tg(a) ≈ 0.75), Sad ≈ (-0.95; 0.5) (tg(a) ≈ -0.53), Angry ≈ (-1; 0.75) (tg(a) ≈ -0.75), Surprised ≈ (0; 0.75) (tg(a) → ∞), Scared ≈ (-0.55; 0.75) (tg(a) ≈ -1.36), Disgusted ≈ (-1; 0.6) (tg(a) ≈ -0.6). Заметим, что это расположение (Рисунок 3.1) несколько отличается от общепризнанного (Рисунок 1.1). Так, Sad должен располагаться в 3-ей четверти координатной плоскости, но из данных следует, что он расположен во 2-й. Можно предположить, почему Arousal ∈ [0; 1], а не [-1; 1]. Скорее всего, значения y-й координаты взяты по модулю. Эмоции Scared и Angry поменялись местами, а Surprised находится значительно левее (на оси y), чем должна.

N //из 1.4 Стоит отметить, что классическое измерение pleasure (удовольствие) из модели PAD в логах представлено как valence (валентность). При этом диапазон допустимых значений никак не меняется и остаётся прежним [-1; 1]. Но при этом координата Arousal находится в диапазоне [0; 1], что нетипично ввиду возможности наличия отрицательных значений Arousal (Рисунок 1.1). D – Dominance измерение опущено ввиду отсутствия необходимых данных.

Координаты семимерного вектора эмоций имеют диапазон [0; 1], хотя в сумме не дают 1. Т.е. возможны случаи, когда испытуемый одновременно демонстрирует несколько эмоций. Тогда сумма всех координат вектора превышает 1.

Трудностями в ходе решения поставленной задачи отображения из двухмерного пространства в семимерное могут стать ситуации, когда испытуемый проявляет несколько эмоций одновременно. В таком случае, 7-ми мерное пространство точнее отображает ситуацию, нежели 2-х мерное. Но поскольку в текущей работе рассматривается отображение из 2-х мерного в 7-ми мерное, то данный случай становится вовсе не совсем корректным т.к. мы пытаемся из менее точной информации получить более точную. Такие ситуации не являются однозначными и могут иметь различные интерпретации в семимерном пространстве для близких значений PA. N

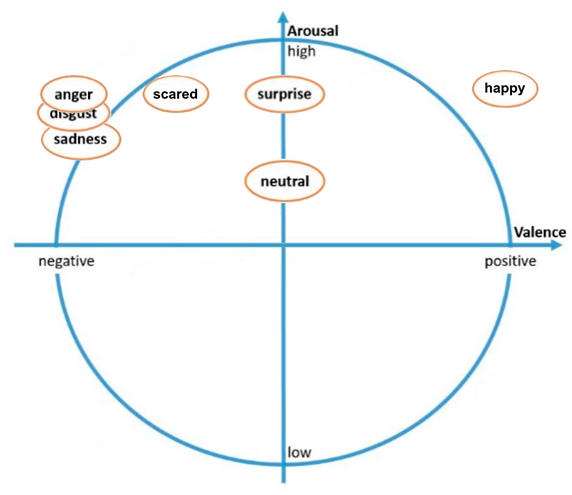


Рисунок 2.1 – Расположения семи чистых эмоций на плоскости PA (VA)

Можно предположить, что нейросетям будет довольно трудно различать близкие по расположению эмоции. Например, Angry, Disgusted, Sad или Neutral, Surprised. Это означает, что значимым критерием оценки корректности обученности нейросети будет являться верное определение эмоции. Нейросеть должна быть чувствительна к нужной эмоции, стараясь как можно меньше путать её с близкими по расположению.

Наиболее корректным критерием для достижения чувстительности нейросети к отдельным схожим эмоциям, кажется способность нейросети чувстовать яркую (чистую) эмоцию. Так, для семимерного вектора, чистой эмоцией будем называть ту, координата которой близка к 1, в то время как значения остальных — близки к 0. Например, чистую эмоцию Surprised можно наблюдать в файле "Participant 1\_Zaharova\_Analysis 5\_video\_20190606\_114933\_detailed.csv" на таймкоде 00:00:06.000. Её координаты (Neutral; Happy; Sad; Angry; Surprised; Scared; Disgusted): (0.035933; 0.001227; 0.000002; 0.000004; 0.985002; 0.001629; 0.000011). Чистая Neutral эмоция присутствует в файле "Participant 1\_Provotorov\_Analysis 10\_video\_20190606\_125247\_detailed.csv" на таймкоде 00:00:09.900. Семимерные координаты соответственно: (0.993989; 0.000053; 0.000353; 0.000277; 0.003347; 0.010323). Аналогично можно найти и другие чистые эмоции (Happy, Sad, Angry, Scared, Disgusted).

Большинство чистых эмоций представлены в достаточном количестве различных экземпляров векторов исходных данных (в логах испытуемых). Трудности возникают лишь с представителями чистого Sad (всего 14 векторов), Scared (всего 12 векторов), Angry (всего 44 вектора). К тому же, Scared имеет max-е значение 0.55, а не 1, что дополнительно усложняет обучение.

Как было сказано ранее в пункте 1.2, помимо распознавания чистых эмоций, трудными случаями считаются ситуации, когда испытуемый демонстрирует несколько эмоций одновременно. В таких векторах заметна неоднозначность интерпретации. Например, в файле "Participant 1\_Aleksashina\_Analysis 1\_video\_20190607\_114857\_detailed.csv" на таймкоде 00:00:38.000 выражение лица испытуемой показывает эмоцию, которая в двухмерном пространстве имеет координаты (0.015823; 0.273556). Анализ данных выявил, что этой точке наиболее вероятно соответствует нейтральная (Neutral) эмоция (0; 0.25), однако FaceReader выдаёт значение семимерного вектора (0.229891; 0.565980; 0.000106; 0.550158; 0.000208; 0.000136; 0.000998). Т.е. наиболее яркими здесь считаются эмоции (Happy (0.565980) и Angry (0.550158)). Этот случай не поддаётся разумному объяснению на основе 2-х мерного пространства. Скорее всего, такой результат получен из данных ещё большей размерности (FACS) где присутствуют 42 координаты. Подобные этому случаи могут служить трудностью при распознавании нейросетью, они противоречивы и их недостаточно, чтобы сформулировать какой-либо законченный вывод.

На основе вышесказанного, можно подтвердить основной критерий для оценивания точности и адекватности модели: научиться распознавать, в первую очередь, чистые эмоции. Требовать от нейросети распознавание заведомо некорректных случаев, типа смеси Happy, Angry, Neutral — не целесообразно. Этот критерий отбора нейросети по правильному установлению чистых эмоций условно назовём «чувстовать нотки».

Критерий «чувстовать нотки» заставляет нас позаботится о сборе чистых эмоций в отдельности и добавления в обучающую выборку. Причём добавить в обучение стоит все, которые сможем найти, т.к. они лишены ошибочной интерпретации смесей нескольких эмоций.

### 2.2.3 Формирование обучающей и тестовой выборок.

Исходные данные необходимо почистить перед тем как добавлять в обучающую выборку. Для избавления от шумов, искусственно обобщим данные. Для этого объединённые данные (> 25 тыс. векторов) сгруппируем по значениям 2-х полей (Valence, Arousal), предварительно округлив каждый из них до 2-х знаков после запятой. Эта дискретизация пар (V; A) делается для того, чтобы увидеть общую закономерность и попробовать построить функциональную зависимость между парой (V; A) и вектором (Neutral; Happy; Sad; Angry; Surprised; Scared; Disgusted).

Таким образом, для всех исходных данных (> 25 тыс. векторов) было сформировано 5066 групп. В каждой группе (паре (V; A) с точностью %.2f) для каждого столбца эмоции вычислим среднее (*mean()*) значение. Таким образом и получим функциональную зависимость между парой (V; A) и семимерным вектором (Neutral; Happy; Sad; Angry; Surprised; Scared; Disgusted).

Каждая группа содержит некоторое количество строк исходных данных. Для сохранения частотных свойств встречаемости строк в группу используем формулу: *int(round(ln(len()), 0)) + 1*. Где ln — логарифм по основанию 10, *len()* - количество строк в группе, *round* — округление до целого (0), а *int* — приведение к целому типу. Например, в группе c (V; A) (-1.0, 0.63) (Disgusted) находятся 5 векторов исходных данных. Тогда в обучающую выборку попадут *int(round(ln(5)*, *0)) + 1* = *int(round(1.61, 0)) + 1* = *int(2.0) + 1* = *3* вектора из группы (со средними (*mean()*) значениями столбцов). Другой пример - в группе c (V; A) (-0.0, 0.25) (Neutral) сосредоточено 1072 строки. Это означает, что в обучающую выборку попадут *int(round(ln(1072), 0)) + 1* = *int(round(6.98, 0)) + 1* = *int(7.0) + 1* = *8* векторов.

Всего таким способом получается получить 6481 строку очищенных данных (на 5066 групп). Эти данные хранятся в *DataFram*'е *df\_train*. Подобная группировка вынесена в отдельную функцию *groupby* модуля *tools* и подробнее описана в пункте 4.2.

Помимо очищенных и сгруппированных данных в обучающую выборку нужно добавить чистые эмоции (7-ми мерные вектора с значениями одной координаты близкой к 1, а остальных — к 0). Их лучше отобрать заранее и сохранить в отдельный файл. Работа с чистыми данными происходит в *DataFram*'е *df\_clear*. Всего чистых эмоций 530 строк, и они также очищенны от шумов аналогично функции *groupby* (но берутся все чистые экземпляры, без формулы *int(round(ln(len()), 0)) + 1)*.

Также в обучающую выборку стоит добавить оставшиеся после группировки данные (например, для группы из 1072 строк — было отобрано 8 и осталось 1064 строки). Всего остаётся чуть меньше 19 тыс. строк после применения функции *groupby*. Они записываются в переменную *df\_other*.

Для сбалансированного обучения нужно взять % оставшихся данных (из *df\_other*) и объединить вместе с *df\_train* и *df\_clear*. Перемешать объединёные данные для обучения помогает реализованный метод *train\_test\_split* (см. пункт 4.2).

## 2.3 Разработка алгоритма обучения моделей.

В данном разделе разрабатываются алгоритмы обучения нейросетевых и регрессионных моделей.

# **Раздел 3. Инженерная часть**

## 3.1 Проектирование архитектуры приложения. Описание архитектуры приложения.

В данном разделе описывается процесс проектирования архитектуры приложения.

## 3.2 Проектирование макетов клиентской части приложения.

В данном разделе происходит проектирование макетов клиентской части приложения.

## 3.3 Проектирование серверной части web-приложения.

В данном разделе осуществляется проектирование серверной части web-приложения.

# **Раздел 4. Технологическая и практическая часть**

## 4.1 Обучение моделей на обучающей и тестирование на тестовой выборках.

В данном разделе производится обучение моделей и тестирование их работоспособности.

## 4.2 Оценка точности работы моделей.

В данном разделе оценивается точность моделей, обученных в разделе 4.1.

## 4.3 Выбор стека технологий для реализации приложения.

В данном разделе выбираются стек технологий и инструменты для реализации приложения.

## 4.4 Реализация клиентской и серверной частей web-сервиса.

В данном разделе представлены программные блоки реализации web-сервиса.

# **Заключение**

В данной работе ведётся разработка web-сервиса, служащего удобным программным интерфейсом для работы с обученными регрессионными и нейросетевыми моделями.

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. МНОГОМЕРНАЯ И ДИСКРЕТНАЯ МОДЕЛИ ЭМОЦИЙ [Электронный ресурс] / URL: https://ozlib.com/851130/psihologiya/mnogomernaya\_diskretnaya\_modeli\_emotsiy
2. Модель эмоционального состояния PAD [Электронный ресурс] / URL: https://ru.abcdef.wiki/wiki/PAD\_emotional\_state\_model
3. Система кодирования движений лица (FACS) - Визуальное руководство [Электронный ресурс] / URL: https://imotions.com/blog/facial-action-coding-system/
4. Система кодирования лицевых движений [Электронный ресурс] / URL: https://ru.wikipedia.org/wiki/Система\_кодирования\_лицевых\_движений
5. Экман, Пол - Критика [Электронный ресурс] / URL: https://ru.wikipedia.org/wiki/Экман,\_Пол#Критика
6. Виртуальный ассистент [Электронный ресурс] / URL: https://ru.wikipedia.org/wiki/Виртуальный\_ассистент
7. Распознавание речи - История [Электронный ресурс] / URL: https://ru.wikipedia.org/wiki/Распознавание\_речи#История
8. Обзор технологий синтеза речи [Электронный ресурс] / URL: https://habr.com/ru/company/tinkoff/blog/474782/
9. Основы построения интеллектуальных систем: учеб. пособ./ Г.В. Рыбина. – М.: Финансы и статистика; ИНФРА-М, 2010. – 432 с.: ил.
10. Цифровые аватары: как виртуальные помощники [Электронный ресурс] / URL: https://vc.ru/future/218083-cifrovye-avatary-kak-virtualnye-pomoshchniki-pereselyayutsya-iz-fantasticheskih-filmov-v-nashi-kvartiry-i-ofisy
11. EMOTION ANALYSIS FaceReader [Электронный ресурс] / URL: https://www.noldus.com/facereader
12. A Circumplex Model of Affect // Journal of Personality and Social Psychology 39(6), 1980, pp. 1161-1178
13. Регрессионный анализ [Электронный ресурс] / URL: https://vc.ru/u/425321-sf-education/224225-chto-takoe-regressionnyy-analiz
14. 5 видов регрессии и их свойства [Электронный ресурс] / URL: https://medium.com/nuances-of-programming/5-видов-регрессии-и-их-свойства-f1bb867aebcb
15. Автокодировщик [Электронный ресурс] / URL: https://ru.wikipedia.org/wiki/Автокодировщик
16. Снижение размерности [Электронный ресурс] / URL: https://ru.wikipedia.org/wiki/Снижение\_размерности#Ядерный\_метод\_главных\_компонент\_(ЯМГК)
17. Экман, Пол. Психология эмоций [= Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life] / Пер. с англ.: В. Кузин. — СПб.: Питер, 2010. — 336 с.
18. Искусственные нейронные сети и приложения: учеб. пособие / Ф.М. Гафаров, А.Ф. Галимянов. – Казань: Изд-во Казан. ун-та, 2018. – 121 с.
19. Искусственный интеллект с примерами на Python. : Пер. с англ. -СПб. : ООО "Диалектика", 2019. -448 с.
20. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2, 3-е изд.: Пер. с англ. СПб. : ООО "Диалектика" 2020. - 848 с.