

На правах рукописи



Макаров Виктор Витальевич

**Методы и алгоритмы автоматической
классификации психофизиологических
характеристик человека**

05.13.17 – Теоретические основы информатики

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Москва – 2022

Работа выполнена в *Федеральном исследовательском центре «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН)*.

Научный руководитель: **Цурков Владимир Иванович**
д.ф.-м.н., профессор,
ФИЦ ИУ РАН,
заведующий отделом №31 (сложных систем)

Официальные оппоненты: **Гостев Иван Михайлович**
доктор технических наук,
ФГБУН ИППИ РАН,
ведущий научный сотрудник

Одиноких Глеб Андреевич
кандидат технических наук,
ООО «Техкомпания Хуавэй»,
руководитель подразделения HiSilicon Kirin AI

Ведущая организация: ФГБОУ ВО Владимирский государственный
университет им. А.Г. и Н.Г. Столетовых.

Защита состоится *«15» сентября 2022 г. в 15:00* на заседании диссертационного совета *Д 002.073.05* при *Федеральном исследовательском центре «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН)* по адресу:
119333, г.Москва, ул.Вавилова, д.40

С диссертацией можно ознакомиться в библиотеке *Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН)* и на сайте <http://www.frcscs.ru>.

Автореферат разослан «_____» 2022 г.

Ученый секретарь
диссертационного совета Д 002.073.05,
кандидат технических наук



Рейер И.А.

Общая характеристика работы

Актуальность темы исследования.

Несмотря на постоянное развитие систем безопасности, человеческий фактор всё ещё остается одной из главных уязвимостей. Использование полиграфа позволяет уменьшить такие риски. Психофизиологические исследования решают следующие прикладные задачи: выявление негативных проявлений в прошлом опыте кандидатов на должность, проверка на лояльность и соблюдение внутренних регламентов организации, осуществление оперативно-розыскной деятельности, проведение корпоративных и антикоррупционных расследований и т.д.

Полиграфная проверка является эффективным, но достаточно трудозатратным и требовательным к квалификации специалиста способом выявления скрываемой информации. Именно поэтому одной из основных целей данной работы является создание системы для автоматических рекомендаций полиграфологу. «Второе мнение» поможет оперативно принять решение или скорректировать саму процедуру полиграфной проверки.

Распознавание эмоций человека является важной научно-исследовательской проблемой, близкой к таким областям, как медицина и психология.

Распознавание эмоций используется в следующих сферах деятельности: онлайн-обучение – построение учебного плана с учётом динамики вовлеченности учащихся на каждом этапе; банковское дело – дополнение скоринговых моделей информацией о возможных искажениях для выявления мошенников; колл-центры – управление удовлетворённостью во время звонка, составление независимого индекса потребительской лояльности; организация транспортной безопасности - контроль за состоянием водителя, сигнализирование о возможном переутомлении; производство эмпатичных роботов – дополнительный инструмент для выбора оттенков диалога.

Цели и задачи диссертационной работы:

В работе были поставлены следующие **цели**:

- Повысить точность методов и алгоритмов классификации эмоционального состояния человека по голосовой фонограмме.
- Разработать методы и алгоритмы для классификации эмоционального состояния человека по видеозаписи, устойчивые к изменениям условий съёмки.
- Создать методы и алгоритмы классификации силы реакции организма на стимулы при помощи регистрируемых полиграфом параметров (дыхание, сердечно-сосудистая и электродермальная активность).

Для достижения поставленных целей были решены следующие **задачи**:

- Создание, исследование и подбор алгоритмов обработки голосовых фонограмм для классификации эмоционального состояния говорящего;
- Исследование и разработка методов классификации эмоций человека по видеозаписям;
- Нормализация психофизиологических характеристик, учитывающих индивидуальные особенности испытуемого;
- Разработка методов и алгоритмов автоматической классификации параметров, регистрируемых при помощи полиграфа – КГР, плеизмограмма, дыхательные циклы;
- Создание тестового приложения и проведение вычислительных экспериментов по определению работоспособности перечисленных методов.

Научная новизна диссертационной работы состоит в следующем:

1. Создана архитектура нейронной сети для автоматической классификации голосовых фонограмм с высокой точностью;

2. Созданы новые методы классификации эмоций по видео, отличающиеся высокой устойчивостью при работе с материалами, записанными в нестацийных условиях;
3. Предложен метод нормализации психофизиологических характеристик, полученных при помощи полиграфа, учитывающий индивидуальные особенности испытуемого.
4. Созданы три новых метода автоматической классификации силы реакции человека (балльная оценка) на предъявляемый стимул при помощи оценки регистрируемых независимых параметров: дыхательных циклов, электрической активности кожи (КГР), сердечных ритмов (плетизмограммы).

Теоретическая и практическая значимость.

Результаты, изложенные в диссертации, применены для создания интеллектуальной системы оценки факторов риска при трудоустройстве и проведения служебных опросов.

Получено 2 свидетельства о государственной регистрации программы для ЭВМ:

1. №2021615620, «Программное обеспечение по оценке эмоционального состояния человека по видеопотоку в режиме реального времени с использованием искусственного интеллекта»;
2. №2022661019 «Программный комплекс на основе инновационной стандартизированной и валидизированной методики для проведения полиграфных проверок»;

Результаты работы реализованы и используются в следующих системах:

- Система автоматической балльной оценки проведенного тестирования в профессиональном компьютерном полиграфе «Финист»;

- Модуль по оценке эмоционального состояния собеседника в аппаратно-программном комплексе «Детектрон».

ФГБУ «Фонд содействия развитию малых форм предприятий в научно-технической сфере» подтвердил практическую значимость указанных выше систем и обеспечил поддержку развития указанных выше модулей в форме грантов: №3493ГС1/57463 «Разработка прототипа программного обеспечения по оценке эмоционального состояния человека по видеопотоку в режиме реального времени с использованием программного обеспечения искусственного интеллекта» - 2020 г.; №240ГС1ЦТС10-D5/65720 «Разработка прототипа программного комплекса на основе инновационной стандартизированной и валидизированной методики для проведения полиграфных проверок» - 2021 г.

Методология и методы исследования. В работе применялись следующие методы анализа аудиосигналов: разложение на эмпирические моды и их ансамбль, мел-кепстральные характеристики, дифференциальный энергетический оператор и дополненный вариант Тигера-Кайзера; вейвлет-преобразования. При анализе видеоизображения использовался метод локальных бинарных паттернов. Для создания нейронных сетей использовались следующие архитектуры: сверточные нейронные сети, многослойный перцептрон, трансформер; классификаторы гауссовских процессов, экстремально рандомизированных деревьев, леса случайных решений, случайного процесса Гаусса, дерева решений, алгоритмы Adaboost, модель смеси Гауссовых распределений. Для визуализации данных высокой размерности применялся метод стохастического вложения соседей с t-распределением.

Положения, выносимые на защиту:

- Предложена и программно реализована архитектура нейронной сети для решения задачи определения эмоции по голосовой фонограмме с высокой точностью.
- Предложена и программно реализована архитектура многослойной ней-

ронной сети для решения задачи определения эмоции человека по видеозаписи, зарегистрированной в нестудийных условиях.

- Предложен метод нормализации психофизиологических характеристик, полученных при помощи полиграфа, учитывающий индивидуальные особенности испытуемого.
- Создан модуль автоматической классификации силы реакции человека на предъявляемые стимулы при помощи оценки регистрируемых полиграфом параметров (дыхание, сердечно-сосудистая и электродермальная активность).

Степень достоверности и апробация результатов.

Достоверность результатов подтверждена экспериментальной проверкой результатов предлагаемых методов на реальных данных, в том числе сторонними организациями; публикациями результатов исследования в рецензируемых научных изданиях и трудах конференций; воспроизводимостью результатов исследования при использовании различных тестовых наборов данных; корректным использованием математического аппарата, стандартных мер качества классификации, современных средств программирования и библиотек машинного обучения.

Основные результаты диссертации докладывались на следующих конференциях: II Всероссийская научная конференция с международным участием «От идеи – к практике: социогуманитарное знание в цифровой среде» - Новосибирск, 2022; 14-я международная научная конференция студентов и магистрантов «Современный специалист-профессионал: теория и практика» - Барнаул, 2022; Всероссийская научная конференция молодых ученых, посвященная Году науки и технологий в России «Наука. Технологии. Инновации» - Новосибирск, 2021.

Публикации. Материалы диссертации опубликованы в 3 печатных работах, из них 3 статьи в рецензируемых журналах из списка ВАК.

Личный вклад автора. Содержание диссертации и основные положения, выносимые на защиту, отражают персональный вклад автора в опубликованные работы. Подготовка к публикации полученных результатов проводилась совместно с соавторами, причём вклад диссертанта был определяющим. Все представленные в диссертации результаты получены лично автором.

Структура и объём диссертации. Диссертация состоит из введения, 3 глав, заключения и библиографии. Общий объём диссертации 104 страницы, из них 86 страниц текста, включая 14 рисунков. Библиография включает 159 наименований на 16 страницах.

Содержание работы

Во Введении обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, рассмотрена история развития метода психофизиологических исследований, показана практическая значимость полученных результатов, представлены выносимые на защиту научные положения.

В первой главе описаны методы распознавания эмоций по голосовой фонограмме для следующей задачи классификации:

X - множество голосовых фонограмм x ;

$Y = \{\text{Спокойствие, Радость, Грусть, Злость, Страх, Удивление, Отвращение}\}$ - множество классов, представляющих собой семь эмоций.

$\{(x_1, y_1), \dots, (x_m, y_m)\}$ - обучающая выборка, где m - размер набора данных

Требуется построить алгоритм $a : X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$, то есть ответить на вопрос «Какую эмоцию испытывал человек на голосовой фонограмме x ?»

Точность классификации модели с учетом сбалансированности набора дан-

ных вычисляется следующим образом:

$$accuracy = \frac{\sum_i TP_i}{m_{test}} \quad (1)$$

где TP_i – количество корректных предсказаний объектов из класса i , m_{test} – размер тестовой выборки.

Для обучения был выбран набор данных RAVDESS, представляющий из себя 1440 аудиозаписей, длительностью по 4 секунды каждого. Класс «Нейтральное состояние» был исключен из обучающей выборки по причине низкой информативности и схожести с классом «Спокойствие».

Свёрточная нейронная сеть для классификации эмоций по голосовым фрагментам.

При помощи вспомогательных библиотек librosa и matplotlib производится первичная обработка файлов: приведение аудиоматериалов к моноформату, подготовка изображения с мел-спектрограммой аудиозаписи с разрешением 640 на 480 пикселей. Посредством библиотеки Keras производится загрузка, нормализация, разделение на обучающую и тестовую выборки, обучение и выбор наилучших архитектур, составление матрицы ошибок и классификационного отчета по распознаванию эмоций.

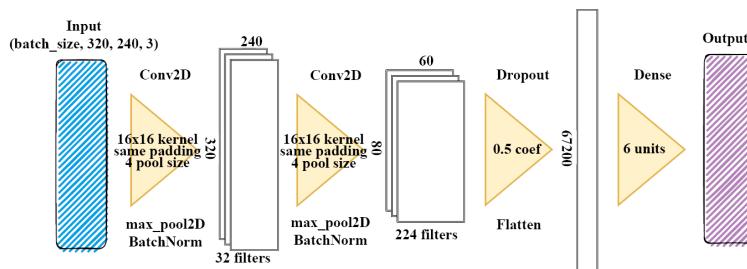


Рис. 1. Архитектура нейронной сети для работы с мел-спектрограммами

Для повышения точности классификации были использованы следующие методы: комбинаций увеличения блоков и свёрточных слоёв, нормализация пакета данных для одной итерации (батчей), тюнинг гиперпараметров с помощью алгоритма RandomSearch. В результате были получены следующие результа-

ты точности классификации для различных комбинаций блоков и свёрточных слоёв: 1conv_2blocks – 80.2%, 2conv_2blocks – 78.29%, 1conv_3blocks – 72.56%, 2conv_3blocks – 78.81%.

Наилучшего результата достигла имеет сеть с архитектурой, представленной на рис. 1:

- входной слой (Input),
- Блок №1: свёрточный слой, слоя нормализации пакетов (Batch Normalization), функция активации ReLU,
- Блок №2: свёрточный слой, слоя нормализации пакетов (Batch Normalization), функция активации ReLU,
- слой выброса (Dropout),
- выравнивающий слой (Flatten),
- полносвязный слой (Dense) с функцией активации softmax.

По результатам обучения отмечено следующее: свёрточные сети, имеющие в своей архитектуре один свёрточный слой, показали более высокую точность классификации по сравнению с результатами сетей, имеющих два свёрточных слоя;

Таблица 1. Итоговая точность классификации эмоций по голосовым фонограммам с применением различных методов

Метод	Точность классификации
Метод эмпирических мод	74,6%
Вейвлет-анализ	72,7 %
Предложенная архитектура	80,2 %

Точность классификации эмоций по голосовым фонограммам при помощи описанных в первой главе методов представлена в таблице 1:

Результаты первой главы опубликованы в работе [1].

Во второй главе рассмотрены методы распознавания эмоций по видеозаписи для решения задачи классификации, схожей с описанной ранее. Но в данном случае присутствуют следующие отличия:

X - множество видеозаписей x ;

$x_i = \{x_{i1}, \dots, x_{il}\}$ - видеозапись, состоящая из кадров x_{ij} , где l - длительность видеозаписи в кадрах,

x_{ij} - матрица целых чисел размера $w \times h \times c$, где w - ширина кадра, h - высота кадра, c - цветность кадра ($c = 3$ для цветного, $c = 1$ для черно-белого);

Для распознавания эмоций по видеозаписям использовались наборы данных RAVDESS (описан ранее) и Aff-Wild.

Набор данных Aff-Wild состоит из 298 видеороликов, общей продолжительностью более 30 часов, на которых зарегистрированы спонтанные реакции человека на различные стимулы (раздражители) в произвольные условия записи. Каждый ролик аннотирован по шкале валентности (позитив-негатив) и возбуждения (расслабленность - тонус). Значения каждого показателя изменяются -1 до 1.

Синтез нового набора данных Aff-Wild-Em (Рис. 2) был сделан посредством следующего сопоставления:

$c_i \in$ Радость (Happy) для $(0.3 < x \leq 1)$ и $(0.3 < y \leq 1)$

$c_i \in$ Спокойствие (Calm) для $(0.3 < x \leq 1)$ и $(-1 \leq y < -0.3)$

$c_i \in$ Грусть (Sad) для $(-1 \leq x < -0.3)$ и $(-1 \leq y < -0.3)$

$c_i \in$ Страх (Scare) для $(-1 \leq x < -0.3)$ и $(0.3 < y \leq 1)$

$c_i \in$ Нейтральность (Neutral) для $(-0.3 \leq x \leq 0.3, x \neq 0)$ и $(-0.3 \leq y \leq 0.3, y \neq 0)$

При удалении кадров с нулевыми метками валентности и возбуждения набор данных становится более сбалансированным по причине их большого ко-

личества и неинформативности.

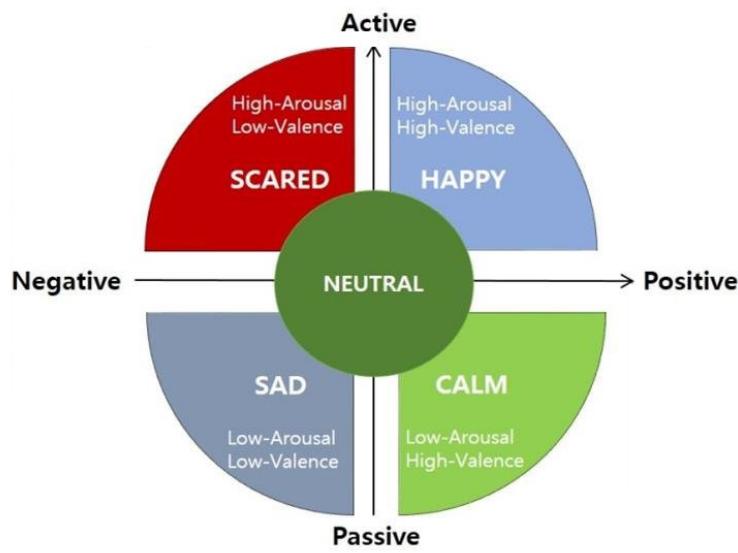


Рис. 2. Схема сопоставления классов для синтеза нового набора данных

Видео читается покадрово при помощи библиотеки Scikit-video. Для удобства работы написан итератор, который определяет характеристики видео из заголовка и затем читает видео поsekундно. Лишние кадры удаляются, а на оставшихся итератор находит лицо при помощи библиотеки Dlib и изменяет его размер до 256x256. Полученные изображения объединяются в партию и обрабатываются нейросетью.

Производится обучение первичной сети на задаче классификации эмоций из нового набора данных (нейтральное состояние, возбужденное позитивное - радость, возбужденное негативное - страх, расслабленное позитивное - спокойствие, расслабленное негативное - грусть). Точность классификации описанной модели составила 60.7%.

Из обученной в предыдущем этапе нейросети убираются последние слои. Полученная нейросеть фиксируется. После чего через нее пропускается набор данных RAVDESS и сохраняется найденный промежуточный результат.

Таким образом получаем промежуточный набор данных для обучения нейросетей по определению эмоций. В результате такой комбинации архитектур

Таблица 2. Таблица итоговой точности классификации эмоций по видеозаписях с применением различных методов

	Aff-Wild-Em	RAVDESS
Local Binary Patterns	-	58,7%
Предложенная архитектура	-	69,4%
После смены набора данных	60,7%	91,0 %

нейросети будут достаточно точны при пропускании через них видеозаписей, зарегистрированных в произвольных условиях. Первичная нейросеть уже умеет с ними работать, а студийные условия выступают в качестве частного случая.

Полноценно архитектура сети выглядит следующим образом:

- нормализация внутри партии и обнуление (dropout) с вероятностью 0.1;
- свёртка с ядром 5x5 и шагом 2;
- 8 плотных свёрток с ядром 3x3 и 15 каналами;
- укрупнение размера промежуточного изображения;
- слой нормализации, слой обнуления;
- свёртки 1x1 с числом каналов, равным половине входящих;

Последний блок, который впоследствии будет отбрасываться:

- обнуление с вероятностью 0.3;
- свёртка размером 90x1x1;
- 6 плотных свёрток с ядром 1x1;
- усреднение слоёв;
- нормализация по партии;
- усреднение каналов до 5 созданных классов.

В отличие от стандартной архитектуры, здесь удалены полно связные слои, образующие итоговый набор классов. Вместо них при достижении слоем размеров менее 5x5 происходит усреднение внутри слоёв.

Точность классификации эмоций по видеозаписях при помощи описанных

во второй главе методов представлена в таблице 2:

Результаты второй главы опубликованы в работе [2].

В третьей главе рассмотрены методы классификации в исследованиях на полиграфе. Применение такого оборудования требует высокой квалификации специалиста, но в то же время нет аналогов по точности и удобству применения в кадровых проверках и различных расследованиях. Полиграф является медико-биологическим прибором для одновременной регистрации следующих психофизиологических характеристик.

- Динамика дыхательных циклов;
- Электрокожное сопротивление (кожно-гальваническая реакция);
- Артериальное давление;
- Периферическое крово направление (фотоплетизмограмма);
- Двигательная активность (тремор).

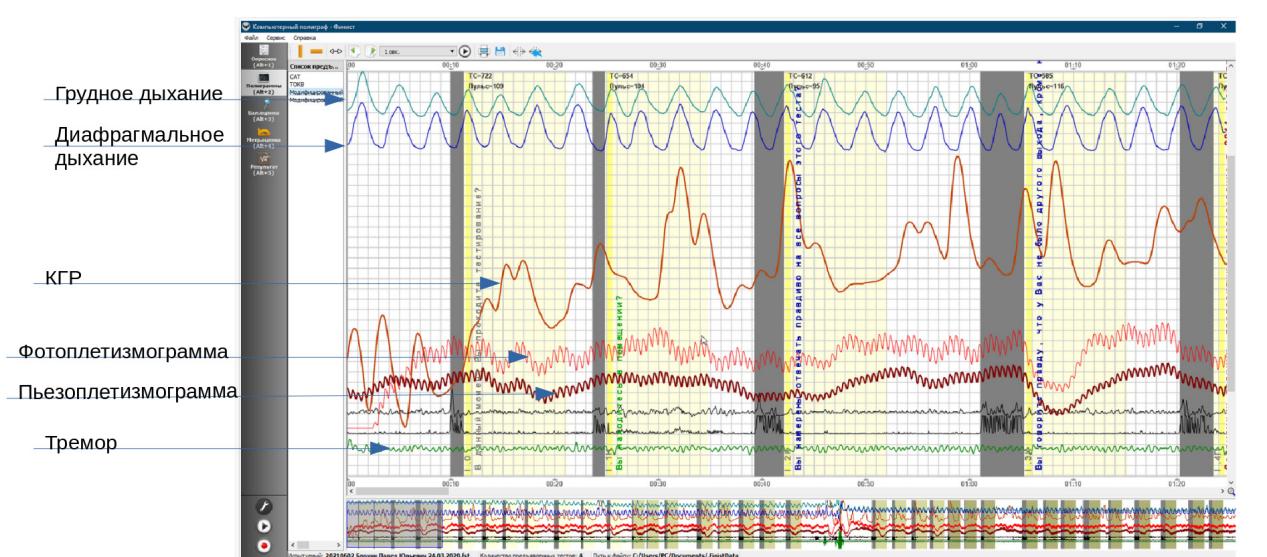


Рис. 3. Скриншот полиграммы в интерфейсе профессионального компьютерного полиграфа «Финист» с подписанными каналами съема психофизиологических характеристик.

Как известно, полиграф работает так: имеется круг вопросов из некоторой предметной области. К испытуемому индивиду подключаются датчики для вышеуказанных измерений. Принимается ответ: либо «да», либо «нет». Все вопросы повторяются не менее трёх раз. Полиграфолог анализирует полученные

результаты с помощью системы трёхбалльной оценки и окончательно констатирует о справедливости ответа.

Суть такой системы заключается в том, что реакциям на каждый вопрос теста по каждому каналу выставляются баллы в следующем порядке: сильные реакции получают по 2 балла, средние - 1 балл, слабые - 0 баллов. Для принятия решения суммируются баллы по 3 каналам по всем повторениям.

Например, если в результате трёх повторений вопроса получилось 11 баллов, то можно говорить о высокой ситуативной значимости реакции и возможном сокрытии информации, если 2 балла – то имеет место слабая реакция (это указывает на то, что индивид не испытывает стресс и ему нечего скрывать).

Для удобства изложения далее балльная оценка реакции заменяется на равнозначный процесс классификации силы реакции на слабую, среднюю и сильную.

На данный момент полиграфологи применяют только прямые методы обсчета полиграмм, для этого вычисляются и сравниваются следующие характеристики:

- длина линии сигнала верхнего (грудного) и нижнего (диафрагмального) дыхания,
- количество дыхательных циклов,
- максимальная амплитуда КГР,
- площадь под графиком сигнала КГР,
- смещение средней линии плеизмограммы и другие.

К сожалению, индивидуальные физиологические особенности каждого человека привносят неравномерные искажения по одному или нескольким сигналам. Прямые методы обсчетов не позволяют сделать автоматическую подстройку к таким изменениям, поэтому точность таких подходов является недостаточной для самостоятельного принятия решения по результатам проведенного

психофизиологического исследования. Далее ставится задача исправить этот недостаток при помощи нейронных сетей и машинного обучения.

Подготовка набора данных из полиграмм.

Рассматриваются 3 следующих независимых задачи классификации:

X_{breath} , X_{kgr} , X_{pg} - множества зарегистрированных полиграфом реакций на вопросы по каналам дыхания (грудного и диафрагmalного), КГР и плеизмограммы (фото и пьезо датчики), состоящие из векторов:

$$x_i = (x_{i1}, \dots, x_{i480}), x_{ij} \in \mathbf{Z} \text{ - для канала дыхания}$$

$$x_i = (x_{i1}, \dots, x_{i240}), x_{ij} \in \mathbf{Z} \text{ - для канала КГР}$$

$$x_i = (x_{i1}, \dots, x_{i480}), x_{ij} \in \mathbf{Z} \text{ - для канала плеизмограммы}$$

$$Y = \{\text{«Слабая реакция», «Средняя реакция», «Сильная реакция»}\};$$

$$\{(x_1, y_1), \dots, (x_m, y_m)\} \text{ - обучающие выборки по каждому каналу}$$

Требуется построить алгоритмы $a : X_{breath} \rightarrow Y$, $a : X_{kgr} \rightarrow Y$, $a : X_{pg} \rightarrow Y$ способные классифицировать произвольный объект $x \in X_{channel}$ (отдельно по каждому каналу), то есть ответить на вопрос: «Насколько сильная реакция x зарегистрирована по выбранному каналу?»

Для обучения автоматической классификации силы реакций из полиграмм были выделены реакции по всем вопросам, а именно 12 секунд сигналов с частотой 20 Гц, т.е. 240 точек по каждому каналу. Так было обработано 90 психофизиологических исследований с разными испытуемыми и получено 8 000 отрезков, по каждому из которых группа профессиональных полиграфологов проставляет силу реакции на вопрос (слабая, средняя, сильная).

Сравнительное тестирование архитектур

Рассматриваются встроенные алгоритмы классификации данных библиотеки scikit-learn. При помощи них создаются модели для оценки силы реакций полиграммах (отдельно по каждому из трёх показателей): 1) метод опорных векторов (SVM); 2) классификация гауссовских процессов (GPC); 3) гауссовский наивный байесовский алгоритм (GNB); 4) метод адаптивного бустинга (AdaBoost); 5) метод многослойного перцептрона (MLP); 6) алгоритм гради-

ентного бустинга (GBM); 7) метод решающих деревьев (DecisionTree); 8) метод случайных лесов (RandomForest); 9) алгоритм чрезвычайно рандомизированных деревьев (ExtraTrees). В итоге получено 27 моделей, по одной на каждый показатель и на каждый алгоритм классификаций.

Далее применяется алгоритм VotingClassifier для повышения точности в случаях несогласованности между классификаторами, который присваивает весовые коэффициенты каждому из них. Итоговое решение для трёх моделей вычисляется по следующей формуле:

$$\max(w_1 p_{11} + w_2 p_{21} + w_3 p_{31}, w_1 p_{12} + w_2 p_{22} + w_3 p_{32}, w_1 p_{13} + w_2 p_{23} + w_3 p_{33}) \quad (2)$$

где w_i - весовой коэффициент i-го классификатора, p_{ij} – вероятность принадлежности реакции к классу j, полученная при помощи классификатора i.

Точность классификации силы реакции по трем классам с использованием алгоритма VotingClassifier, включающего в себя три najleпших модели, составила 52%, а из пяти – 59%.

При повторном анализе файлов полиграмм было отмечено, что классы «слабая» и «сильная» реакции, представляющие собой минимальные и максимальные проявления дестабилизации организма соответственно, являются наиболее различимыми и визуально и численно. В то же время среди экземпляров целевого класса «средней» реакции мера различимости мала.

Метод нормализации психофизиологических характеристик, полученных при помощи полиграфа

Структура каждой полиграфной проверки выглядит следующим образом:

Тест №1: предъявление №1, вопросы №1, №2, ..., №12

Тест №1: предъявление №2, вопросы №1, №2, ..., №12

Тест №1: предъявление №3, вопросы №1, №2, ..., №12

(возможный перерыв)

Тест №2: предъявление №1, вопросы №1, №2, ..., №12

Тест №2: предъявление №2, вопросы №1, №2, ..., №12

Тест №2: предъявление №3, вопросы №1, №2, ..., №12

Согласно методологии полиграфных проверок, признаки силы реакции могут меняться от предъявления к предъявлению даже в рамках одного блока вопросов (теста). А индивидуальные особенности испытуемых сохраняются на протяжении всей проверки.

На основании этого был сделан вывод о том, что нормализация должна проходить по наиболее стабильному этапу проверки. Поэтому был применен метод z-масштабирования по каждому предъявлению:

$$z_{ijkl} = \frac{x_{ijkl} - \mu_{ijk}}{\sigma_{ijk}} \quad (3)$$

где i - идентификатор испытуемого (исследования), j - номер теста, k - номер предъявления, l - номер вопроса, μ_{ijk} и σ_{ijk} - среднее значение и стандартное отклонение значений психофизиологических показателей на соответствующем предъявлении.

По результатам проведенных численных экспериментов наиболее высокую точность классификации удалось получить именно с применением описанного выше метода нормализации в рамках одного повторения одного блока вопросов одной полиграммы.

Применение архитектуры трансформера

Для повышения точности был реализован классификатор для двух классов с архитектурой трансформера, которой можно обрабатывать последовательности не по порядку. В классическом варианте он представляет из себя две части – кодировщик и декодировщик.

В данном случае была использована структура, состоящая в части кодировщика из чередующихся слоёв внимания (внимание на основе скалярного произведения) и многослойного перцептрана, а также декодировщика (простого классификатора). Обе части трансформера реализованы стандартными средствами

библиотеки Keras.

Данные на вход подаются в размерности (26; 240), т.е. единовременно подается 26 образцов данных (все возможные каналы полиграфа) размером 240 значений. Затем данные кодируются при помощи одноразмерного (1D) PatchEncoder с шириной (8; 2), т.е. линейно трансформируются проецированием на вектор.

Кодированные данные поступают на группу чередующихся слоёв внимания и перцепtronов, а затем поступают на слой SeqPool для субдискретизации. Этап подвыборки последовательности нормируется (LayerNormalization) средствами библиотеки Keras и данные с него затем поступают на полно связный слой в 600 нейронов с активацией SeLu. Результат работы кодировщика на этом этапе укрупняется слоем FeaturePooling до двух групп.

Главные характеристики:

размер набора данных (batch size): 60,

количество блоков (чередование внимания и перцептрана): 6,

размерность слоя внимания: 312,

выброс слоя внимания: 0.1 (10

функция трансформации ядра: softmax,

количество наборов матриц весов (запросов, ключей, значений): 6.

Текущая архитектура скомпилирована с использованием оптимизатора Adam (коэф. обучения 0.01) и функции потерь SquaredHinge (квадратичная кусочно-линейная функция потерь). Подгонка осуществлялась на протяжении 50 эпох с применением ранней остановки, если валидационная точность не увеличивается на протяжении 12 эпох и уменьшением коэффициента обучения на пяти эпохах.

Для визуализации и оценки результатов работы предложенной архитектуры данные были обработаны нейронной сетью без последнего слоя. Полученные векторы для каждой характеристики нормируются и кластеризуются при помощи метода стохастического вложения соседей с t-распределением (t-SNE), который итеративно размещает объекты в пространстве более низкой размерности с условным сохранением расстояния между соседями.

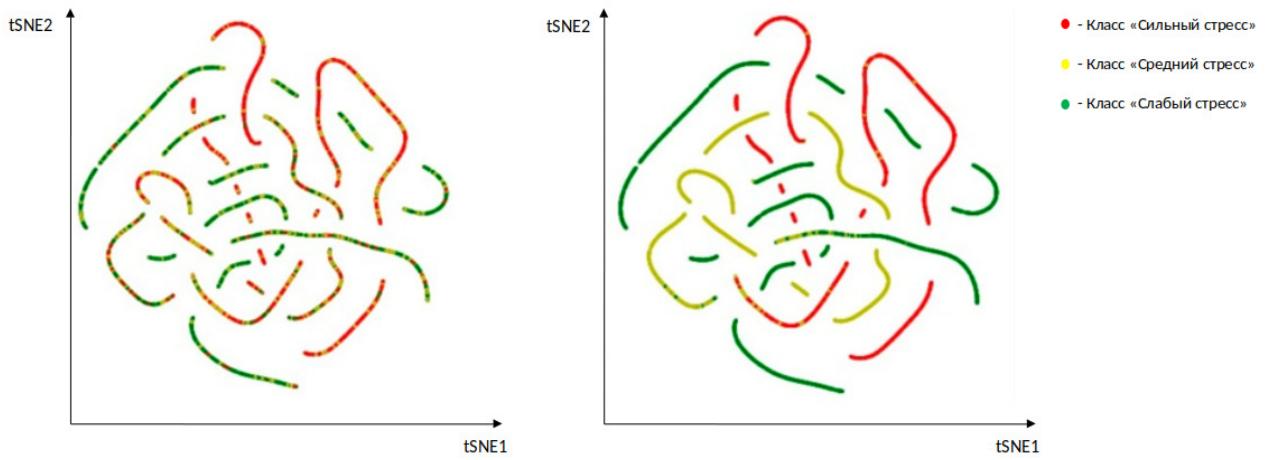


Рис. 4. Результаты кластеризации набора данных по каналу фото и пьезо плеизомограммы методом стохастического вложения соседей с t-распределением (t-SNE): экспертная разметка (слева) и результаты классификации предложенным трансформером (справа).

Далее сравниваются графики с классификацией при помощи экспертной разметки и предложенного трансформера. На Рис. 4 можно заметить, что распознанные машиной точки «Средняя реакция» разбросаны не хаотично, а образуют сплошные непрерывные линии после обработки трансформером. Кроме того, точки категории «Средняя реакция» намного реже попадают в кластеры «Слабая реакция» и «Сильная реакция» по сравнению с ручной разметкой.

По результатам исследования можно сделать следующие выводы:

- Увеличение числа блоков в архитектуре дает умеренную точность (до 4го блока зависимость сильная, после 8го – низкая). В отличии от зависимости точности и количества в оценке изображений.
- Нестабильные признаки образуют большое количество нейронов со значением, близким к нулю. Поэтому метод стохастического градиентного спуска (SGD) показывает высокую эффективность на 1-2 эпохах.
- Для обучения архитектур потребовалось 5 эпох для свёрточных сетей, 20 для трансформера. «Укрупнение» входных данных ускоряет обучение трансформера.

Таблица 3. Точность автоматической классификации силы реакции на полиграммах

Канал регистрации	Voting3, 3 кл.	Voting5, 3 кл.	Трансформер, 2 кл.
Дыхание	45,6%	49,3%	72,7%
КГР	52,1%	59,3%	95,3%
ФПГ	46,9%	53,7%	86,8 %

- Для предотвращения переобучения в обязательном порядке следует использовать методы автоблокировки (понижение скорости через 5 эпох, остановка через 20).

В настоящий момент данные нейронные сети интегрированы в программное обеспечение профессионального компьютерного полиграфа «Финист». Получены результаты по эффективности для оценки факторов риска кандидатов при трудоустройстве.

Точность классификации силы реакции по трем каналам на полиграммах при помощи описанных в третьей главе методов представлена в таблице 3:

Результаты третьей главы опубликованы в работе [3].

Выводы

1. Предложена и программно реализована архитектура нейронной сети для решения задачи определения эмоции по голосовой фонограмме с высокой точностью.
2. Предложена и программно реализована архитектура многослойной нейронной сети для решения задачи определения эмоции человека по видеозаписи, зарегистрированной в нестудийных условиях.
3. Предложен метод нормализации психофизиологических характеристик, полученных при помощи полиграфа, учитывающий индивидуальные особенности испытуемого.
4. Создан модуль для автоматической классификации силы реакции че-

ловека на предъявляемые стимулы при помощи оценки регистрируемых полиграфом параметров (дыхание, сердечно-сосудистая и электродермальная активность).

Публикации соискателя по теме диссертации:

1. Деревягин Л.А., Макаров В.В., Цурков В.И., Яковлев А.Н. Интеллектуальная система для определения эмоций на аудиозаписи с помощью мел-спектрограмм // Изв. РАН ТИСУ. 2022. №3. С. 116-121.
2. Ахияров Ф.Р., Деревягин Л.А., Макаров В.В., Цурков В.И., Яковлев А.Н. Покадровое определение эмоций на видеозаписи с применением многослойных нейронных сетей // Изв. РАН ТИСУ. 2022. №2. С. 80-85.
3. Деревягин Л.А., Макаров В.В., Молчанов А.Ю., Цурков В.И., Яковлев А.Н. Применение нейронных сетей в исследованиях на полиграфе // Изв. РАН ТИСУ. 2022. №4. С. 80-85.