УДК 004.032.26
DOI 10.17513/snt.40045

РАЗРАБОТКА МУЛЬТИМОДАЛЬНОГО МЕТОДА СЕНТИМЕНТ-АНАЛИЗА ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ОРГАНИЗАЦИЯХ

Фазульянов Д.В., Гусева А.И.

Национальный исследовательский ядерный университет «МИФИ», Москва, e-mail: fazulianov.dmitrii@gmail.com, aiguseva@mephi.ru

Данная статья посвящена разработке метода мультимодального сентимент-анализа. Сентимент-анализ, известный также как анализ тональности, традиционно применяется в текстовой форме для выявления настроений в социальных медиа, потребительских отзывах и других областях, где критически важно понимание человеческих эмоций. Однако традиционные одномодальные подходы, ограничивающиеся только текстом, часто не учитывают нюансы, передаваемые интонациями, жестами или мимикой, доступные в аудио- и видеоданных. Разработка мультимодального метода сентимент-анализа открывает новые возможности для систем управления и принятия решений в организациях. Интеграция разнородных данных обеспечивает комплексное понимание человеческих эмоций, что позволяет повысить точность и обоснованность принятия управленческих решений. Использование мультимодального метода способствует более точному определению настроений и предпочтения клиентов, что является одним из ключевых факторов для поддержки принятия решений в области стратегического планирования и управления изменениями. С учетом современных достижений в области компьютерного зрения и обработки естественного языка в данной статье предлагается к рассмотрению комплексная методика, объединяющая текстовые, аудио- и видеоданные для создания полной картины эмоционального состояния. Предложенный подход позволяет не только улучшить точность определения тональности, но и глубже понять ее контекст и вариативность. Представленный подход мультимодального сентимент-анализа демонстрирует, как технологии машинного обучения могут синергетически взаимодействовать для обработки разнородных данных, предоставляя возможности для исследования в области социальных медиа, маркетинга и бизнеса, а включение мультимодального анализа в процессы принятия решений предоставляет организациям механизм для улучшения их стратегической гибкости и оперативной эффективности.

Ключевые слова: сентимент-анализ, анализ тональности, тональность текста, тональность видео, тональность аудио, мультимодальный сентимент-анализ, организационная система, механизмы принятия решений на основе данных

DEVELOPMENT OF A MULTIMODAL METHOD OF SENTIMENT ANALYSIS TO SUPPORT DECISION-MAKING IN ORGANIZATIONS

Fazulianov D.V., Guseva A.I.

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, e-mail: fazulianov.dmitrii@gmail.com, aiguseva@mephi.ru

This article is devoted to the development of a method of multimodal sentiment analysis. Sentiment analysis, also known as sentiment analysis, has traditionally been used in text form to identify moods in social media, consumer reviews, and other areas where understanding human emotions is critically important. However, traditional single-modal approaches, limited only to text, often do not take into account the nuances conveyed by intonation, gestures or facial expressions available in audio and video data. The development of a multimodal method of sentiment analysis opens up new opportunities for management and decision-making systems in organizations. The integration of heterogeneous data provides a comprehensive understanding of human emotions, which improves the accuracy and validity of management decisions. The use of the multimodal method contributes to a more accurate determination of customer attitudes and preferences, which is one of the key factors for strategic planning and change management. Using modern advances in computer vision and natural language processing, this article offers a comprehensive methodology that combines text, audio and video data to create a complete picture of the emotional state. The proposed approach makes it possible not only to improve the accuracy of determining tonality, but also to better understand their context and variability. The presented approach of multimodal sentiment analysis demonstrates how machine learning technologies can work synergetically to process heterogeneous data, providing opportunities for research in the field of social media, marketing and business, and the inclusion of multimodal analysis in decision-making processes provides organizations with a tool to improve their strategic flexibility and operational efficiency.

Keywords: sentiment analysis, text tonality, video tonality, audio tonality, the organizational system, multimodal sentiment analysis, data-based decision-making mechanisms

Сентимент-анализ (или анализ тональности) является классом методов в компьютерной лингвистике для определения эмоциональной окраски в различных типах данных. Сентимент-анализ широко применяется для мониторинга социальных медиа, анализа потребительских отзывов, в маркетинге и в других областях, где важно

понимание чувств и мнений людей по отношению к какому-то событию или понятию, например атомной энергетике [1].

Как правило, при анализе тональности выделяют положительную тональность (благоприятные, утвердительные или оптимистические эмоции), негативную тональность (неблагоприятные или пессимистические эмоции) и нейтральную тональность (отсутствие ярко выраженных эмоциональных оценок или чувств) [2].

Еще одним смежным понятием в сентимент-анализе является эмоциональная окраска. В отличие от общего понятия тональности, эмоциональная окраска позволяет более точно классифицировать специфические эмоции, такие как удивление, страх, печаль, эйфория [3]. Это различие важно в контексте мультимодального сентимент-анализа, где разные каналы (текст, аудио и видео) могут передавать разные аспекты эмоций.

В условиях цифровой трансформации организационные системы сталкиваются с необходимостью быстрого и точного принятия решений. Мультимодальный сентимент-анализ предоставляет эффективный механизм для повышения эффективности таких систем, позволяя управляющим структурам лучше понять и предвидеть реакции клиентов на текущее положение, а также на различные изменения и нововведения. Прежде чем перейти к обсуждению сентимент-анализа как метода математического и компьютерного моделирования, важно кратко рассмотреть процесс выражения эмоций с точки зрения нейробиологии, включая социокультурные аспекты. Само научное изучение эмоций затрагивает множество дисциплин – от нейробиологии и нейрофизиологии, которые изучают механизмы возникновения и распознавания эмоций, до клинической медицины, эволюционной биологии и когнитивистики. В основе научного понимания эмоций лежат следующие аспекты.

1. Нейробиологические основы эмоций – многие медицинские исследования показывают, что такие структуры мозга, как гипоталамус, префронтальная кора и миндалевидное тело, играют ключевую роль в обработке и регуляции эмоций. Эти области мозга отвечают за аверсивные сигналы, например страх и тревогу, и учувствуют в высших когнитивных функциях – принятии решений и эмоциональной регуляции [4].

2. Выражение эмоций через жесты и мимику – работы Пола Экмана начиная с 1967 года демонстрируют универсальность выражений лица для основных эмоций в различных культурах. Пол Экман идентифицировал шесть базовых эмоций (счастье, грусть, страх, отвращение, удивление и гнев) и разработал универсальную систему кодирования действий лица FACS, которая позволяет идентифицировать почти все возможные выражения лица. Этот метод и по сей день используется в методах глубокого обучения для определения векторов по изображениям лиц [5, 6].

3. Когнитивный аспект эмоций – согласно теории Ричарда Лазаруса, эмоции не возникают автоматически в ответ на события, а являются результатом когнитивной оценки, которая включает анализ соответствия событий личным желаниям, важности и возможности эмоциональной адаптации при понимании причины происходящего [7].

4. Влияние культуры на выражение эмоций – хотя основные эмоции универсальны, культурные и социальные нормы могут повлиять на то, как эмоции выражаются и интерпретируются.

Таким образом, одномодальные подходы к сентимент-анализу, например анализ исключительно текстовых данных, сталкиваются с серьезными ограничениями. Текстовые данные хоть и информативны, но не всегда способны передать полный спектр эмоционального фона, который может быть выражен через интонацию, жесты или выражения лица в аудио и видео. Например, иронию и сарказм часто трудно уловить только по тексту, без учета тона и контекста речи. Исследования показывают, что добавление аудиовизуальных данных значительно улучшает точность определения эмоций. Исследуя аудио и видео, можно обнаружить микроэкспрессии, вариации тембра голоса и другие невербальные сигналы, которые не учитываются при анализе только текста [8]. Пренебрежение аудиовизуальным контекстом приводит к упрощению восприятия сообщений, что критично в коммуникации, где звук и изображения играют центральную роль в передаче информации и эмоций.

С учетом этих ограничений одномодальных методов становится очевидной потребность в создании методов, интегрирующих мультимодальные данные, для более полного и точного понимания эмоциональной окраски. Предложенный в статье метод объединяет текстовые, аудио- и видеоданные, что способствует более глубокому анализу и повышает точность сентимент-анализа. Такой подход включает анализ широкого спектра данных – от текста до видео, что помогает создавать объективную картину эмоционального отношения к продукции, услугам или деятельности организации, что, в свою очередь, важно для принятия обоснованных управленческих решений, направленных на повышение удовлетворенности клиентов.

Целью исследования является разработка метода мультимодального сентимент-анализа. Этот метод интегрирует данные из различных каналов: текст, аудио и видео – для создания более глубокого и точного анализа эмоциональной окраски информации. Метод предназначен для интеграции в системы поддержки принятия решений, что не только позволит организационным системам эффективнее реагировать на изменения в настроениях и предпочтениях клиентов, но и обеспечит более обоснованное и стратегическое планирование.

Материал и методы исследования

Данное исследование фокусируется на методах сбора данных из различных источников, включая социальные сети, подкасты, телевизионные программы, а также видео- и аудиозаписи. Используются методы очистки и унификации для каждого типа данных (текст, аудио и видео), что необходимо для подготовки данных к дальнейшему анализу. Применяются специализированные методы и модели для каждого типа данных.

• Для видеоданных используется сверточная нейронная сеть (convolutional neural network, CNN) ResNet-50 для обработки визуальной информации и выявления эмоциональных сигналов.

• Для аудиоданных применяются такие методы, как спектральное преобразование и последующая обработка с использование LSTM (Long-Short-Term-Memory), одной из разновидностей рекуррентных нейронных сетей (англ. recurrent neural network, RNN), для выявления эмоциональных нюансов в речи.

• Для текстовых данных используется трансформер BERT (Bidirectional Encoder Representations from Transformer) для извлечения текстовых признаков, учитывающих контекстуальные связи.

Для классификации эмоциональной окраски был разработан мультимодальный подход, в основе которого лежит агрегация признаков из различных источников. Этот процесс включает в себя обучение модели на основе метода опорных векторов (англ. support vector machine, SVM), классификацию сентиментов и последующую оценку эффективности полученной модели.

Результаты исследования и их обсуждение

Шаг 1: Сбор данных

Сбор данных является первым этапом любого метода сентимент-анализа. Этот процесс включает в себя отбор данных, который может существенно различаться в зависимости от объектов исследования, платформ, на которых они размещены, и конечной цели анализа. Этот этап подразумевает многогранный подход к сбору данных, охватывающий текст, аудио и видео, что позволяет получить комплексное представление о передаваемых эмоциях.

• *Текстовые данные* – в зависимости от задачи текстовые данные могут быть собраны из социальных сетей, агрегаторов новостей, форумов, блогов и транскриптов аудио и видео.

• *Аудиоданные* – сбор аудио может включать, например, аудиодорожки подкастов, интервью и телевизионных программ, где звук является ключевым каналом передачи информации.

• *Видеоданные* – как и аудио, видеоматериалы собираются из различных источников, используются инструменты для извлечения как визуальной информации, так и аудиокомпонента.

Далее каждый собранный набор данных подвергается предварительной обработке и очистке для извлечения релевантной информации.

Шаг 2: Предобработка данных и извлечение признаков

Для эффективного сентимент-анализа важен тщательный процесс предобработки данных, который включает ряд специфических методов и инструментов для каждого типа данных. Этот шаг критичен, так как качество и достоверность входных данных напрямую влияют на результаты анализа.

Предобработка и извлечение признаков из видеоданных

Процесс предобработки и извлечения признаков из видеоряда, представленный на рисунке 1, осуществляется с использованием сверочной нейронной сети ResNet-50, которая может эффективно обрабатывать визуальные признаки, связанные с эмоциональными проявлениями в видео [9].

Изначально видеоряд разделяется на отдельные кадры, которые затем масштабируются до стандартных размеров входа для ResNet-50 и подвергаются нормализации цветовых каналов. Далее каждый кадр подается на вход модели ResNet-50 для получения вектора признаков. Эта модель глубокого обучения была предварительно обучена на большом наборе данных изображений [10] и способна выявлять сложные визуальные паттерны, что делает ее применимой для извлечения эмоциональных сигналов из видеоряда. Следующий шаг – агрегация признаков из отдельных кадров

путем усреднения векторов признаков по всем кадрам:

$$v_{video} = \frac{1}{N}\sum_{i=1}^{N} v_i ,$$

где v_i – вектор признаков i-го кадра, N – количество кадров в видео.

Рис. 1. Предобработка и извлечение признаков из видеоряда с использованием архитектуры сверочной нейронной сети ResNet-50

Предобработка и извлечение признаков из аудиоданных

Аудиоданные требуют отдельного набора техник предобработки и извлечения признаков в целях подготовки звуковой информации для анализа и классификации. На рисунке 2 представлен подход, основанный на использовании рекуррентных архитектур как LSTM [11], которые хорошо подходят для обработки временных последовательностей. Такие сети способны улавливать зависимости в аудиоданных, что важно для распознавания эмоциональных нюансов в речи. Например, повышение тональности и ускорение речи могут сигнализировать о волнении или стрессе, а более медленная и монотонная речь может указывать на грусть или усталость.

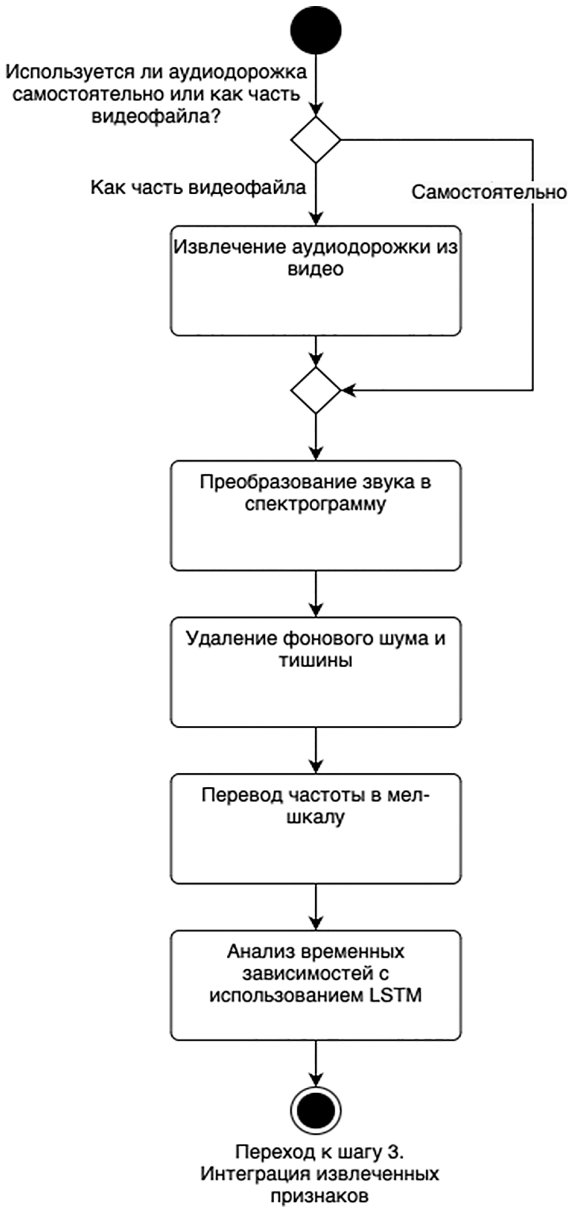

Рис. 2. Предобработка и извлечение признаков из аудиоряда с использованием рекуррентной архитектуры LSTM

Для преобразования звука в спектрограмму используется преобразование Фурье для

перевода временной последовательности аудиосигнала в частотное представление:

$$X(k)=\sum_{n=0}^{N-1}x(n)\times e^{-i2\pi kn/N},$$

где $X(k)$ – спектральные компоненты аудиосигнала, $x(n)$ – амплитуды звуковой волны в дискретные моменты времени, N – общее количество точек в анализируемом фрагменте.

Для большей чувствительности к частотам, близким человеческому уху, частота переводится в мел-шкалу:

$$m=2595\log_{10}\left(1+\frac{f}{700}\right)$$

где m – частота в мелах, число 2595 используется для приближения линейного восприятия частоты звука человеческим ухом в мелах, f – значение частоты в герцах.

Далее (мел-спектрограммы) входные данные подаются на вход LSTM-слоям [12], которые обрабатывают аудиопоследовательность, сохраняя информацию о предыдущих состояниях для определения текущего состояния эмоций:

$$\tilde{n}_t=f_t\times c_{t-1}+i_t\times\tilde{c}_t$$

$$h_t=o_t\times\tanh(c_t),$$

где c_t и h_t – состояние ячейки и скрытое состояние ячейки в момент времени t, f_t, i_t, o_t – вентили «забывания» входа и выхода LSTM.

Эти этапы позволяют извлекать эмоциональные признаки из аудиоряда (v_{audio}), которые затем используются для сентимент-анализа с учетом интеграции между другими типами данных.

Предобработка и извлечение признаков из текстовых данных

Текстовые данные требуют комплексного подхода к предобработке, чтобы преобразовать сырые тексты в формат, пригодный для машинного обучения и анализа. На рисунке 3 представлены основные этапы предобработки текстовых данных и извлечения признаков с помощью модели BERT:

В случае если текстовые данные извлекаются из аудиофайла, то проводится транскрибация с помощью модели Whisper: сначала модель преобразует входящий аудиосигнал в спектрограмму для анализа аудио на уровне частотных компонентов, затем Whisper идентифицирует слова и фразы в аудио, учитывая контекст речи. В конечном итоге, на основе распознанных слов или фраз модель формирует текстовую транскрибацию.

Рис. 3. Процесс предобработки и извлечения признаков из текстовых данных

Далее текстовые данные подвергаются процессу обработки – текст разбивается на токены (на уровне слов), исключаются стоп-слова, которые не несут смысловой нагруз-

ки (союзы, частицы, местоимения, междометия, предлоги, вводные слова и знаки препинания). Для унификации и уменьшения размерности входных данных все слова приводятся к нижнему регистру и приводятся к их словарным формам (леммам).

После процесса предобработки данные подаются на вход модели BERT [13, 14], где она преобразует каждый токен в векторы, кодирующие контекстуальные связи между словами. BERT использует «механизмы внимания», что позволяет сконцентрироваться модели на релевантных словах. Полученные векторы представляют собой эмбеддинги слов, содержащие информацию не только о самом слове, но и о его контексте в рамках текста. Для получения единого векторного представления всего текста используется метод взвешенного усреднения:

$$v_{text} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i w_i,$$

где v – итоговый вектор всего текста, N – количество слов в тексте, α_i – вес, отражающий важность i-го слова, w_i – вектор i-го слова.

Шаг 3: Интеграция извлеченных признаков

Полученные векторные представления из различных типов данных (видео, аудио и текста) комбинируются для создания единого представления. Это позволяет учесть комплексное взаимодействие между разными типами данных, что важно для точного анализа эмоций. Сама интеграция может быть реализована путем взвешенного сложения. Это один из наиболее простых подходов, где каждый набор признаков умножается на заданный вес, который отражает его значимость, и суммируется с другими для получения единого представления:

$$v_s = v_{video} \times w_{video} + v_{audio} \times w_{audio} + v_{text} \times w_{text}$$

Шаг 4: Классификация сентимента

Мультимодальный подход к классификации сентимента основан на использовании метода опорных векторов, который способен эффективно работать с высокоразмерными данными. Этот этап начинается с нормализации данных для выравнивания диапазона признаков, что предотвращает искажение результатов из-за атрибутов с более высокими значениями. Для реализации мультимодального подхода к классификации сентимента были использованы современные программные библиотеки, включая torch, librosa, numpy, OpenCV, sklearn, transformers и whisper. Эти инструменты обеспечили комплексную обработку данных – от извлечения признаков до обучения и валидации моделей.

Модель обучалась и тестировалась на данных из датасета eNTERFACE [15], который содержит большую коллекцию из 1252 эмоциональных выражений 42 актеров в видеоформате на английском языке. Актеры представляют 14 различных национальностей, при этом мужчины составляют 81%, а женщины – 19%.

Для анализа использовались признаки, извлеченные из трех модальностей данных:

• для видеоряда – признаки лица и жестов;

• для аудиоряда – звуковые характеристики, такие как тембр и энергия в частотных диапазонах, извлеченные с помощью мел-кепстральных коэффициентов (англ. mel-frequency cepstrum, MFCC);

• для текста – лингвистические и семантические особенности (контекстуальная связанность, семантическое отношение между словами и синтаксические зависимости), извлеченные с помощью модели BERT.

Чтобы минимизировать риски переобучения, модель использует различные методики, включая регуляризацию в SVM и дропаут в модели BERT при извлечении текстовых признаков. Для проверки устойчивости и обобщающей способности модель применяет кросс-валидацию, что позволяет оценить модель на различных подвыборках данных, тем самым минимизируя риск переобучения и гарантируя более надежную оценку производительности модели. Производительность модели оценивалась с помощью следующих метрик (табл. 1):

• Precision (точность) – доля правильно идентифицированных объектов среди всех объектов.

$$\text{Precision} = \frac{TP}{TP + FP}$$

• Recall (полнота) – отношение верно классифицированных объектов класса к общему числу элементов этого класса.

$$\text{Recall} = \frac{TP}{TP + FN}$$

• Accuracy (общая точность) – доля всех правильно классифицированных случаев.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

• F1-score (F1-мера) – гармоническое среднее между точностью (precision) и полнотой (recall).

$$\text{F1} - \text{score} = 2 \times \frac{precision \times recall}{precision + recall}$$

Таблица 1

Классификационные исходы

	Принадлежит классу (P)	Не принадлежит к классу (N)
Предсказана принадлежность классу	True positive (TP)	False positive (FP)
Предсказано отсутствие принадлежности к классу	False negative (FN)	True negative (TN)

Таблица 2

Сравнительная характеристика мультимодальной модели сентимент-анализа с традиционными одномодальными подходами

Метод	Avg. precision	Avg. recall	Avg. accuracy	Avg. F1-score
Мультимодальная модель SVM	0.92	0.90	0.91	0.91
SVM (текстовые данные)	0.85	0.87	0.86	0.86
Наивный байесовский классификатор (текстовые данные)	0.85	0.86	0.85	0.85
BERT (текстовые данные)	0.83	0.89	0.89	0.86

Эффективность мультимодальной модели сравнивалась с другими популярными методами сентимент-анализа, такими как наивный байесовский классификатор, классификация текстовых данных с помощью SVM и текстовый анализ с использованием BERT. Для обеспечения корректности сравнения все методы, включая мультимодальную модель, наивный байесовский классификатор и анализ с использованием BERT, были обучены и провалидированы на одних и тех же данных. В таблице 2 представлено сравнение усредненных по классам метрик производительности различных методов.

Эти данные демонстрируют, что мультимодальный сентимент-анализ имеет преимущества над традиционными одномодальными подходами, что отражается в повышении всех ключевых метрик производительности.

Заключение

В данной статье представлен метод мультимодального сентимент-анализа, который комбинирует данные из текста, аудио и видео. Такой подход позволяет комплексно анализировать эмоциональный контекст, обогащая анализ за счет использования различных каналов информации. Мультимодальный сентимент-анализ позволяет воспроизвести более полную картину эмоциональных реакций, которые могут остаться незамеченными при использовании традиционных одномодальных подходов.

В настоящей статье были рассмотрены методы глубокого обучения, такие как сверточные и рекуррентные нейронные сети, которые являются эффективными инструментами для извлечения признаков из каждого типа данных. Использование CNN и RNN позволяет обрабатывать и анализировать большие объемы данных, что также важно в контексте обработки мультимодального сентимент-анализа.

В завершение отметим, что использование мультимодального подхода показывает большую эффективность по сравнению с одномодальными подходами и открывает новые возможности для разработки более чувствительных к эмоциональному контексту методов, способствуя повышению точности и скорости принятия решений в организационных системах.

Список литературы

1. Гусева А.И., Кузнецов И.А., Бочкарёв П.В., Смирнов Д.С. Цифровая тень российских международных мегапроектов строительства АЭС за рубежом: оценка тональности высказываний // Современные наукоемкие технологии. 2022. № 1. С. 32-39. DOI: 10.17513/snt.39006.

2. Гусева А.И., Киреев В.С., Бочкарёв П.В. Исследование цифровой тени проектов строительства российских АЭС за рубежом с помощью методов интеллектуального анализа текстов // Приборы и системы. Управление, контроль, диагностика. 2023. № 6. С. 50–57.

3. Indurkhya N., Damerau F. J. Handbook of natural language processing. Chapman and Hall/CRC, 2010. 719 с.

4. Etkin A., Büchel C., Gross J. J. The neural bases of emotion regulation // Nature reviews neuroscience. 2015. Т. 16, №. 11. С. 693-700.

5. Eckman P., Friesen W., Hager, J.C. Facial action coding system (facs): a technique for the measurement of facial action // A8. 1978. Vol. 5. №. 3. P. 56-75.

6. Ramachandran V.S. Encyclopedia of human behavior. Academic Press, 2012. 719 с.

7. Smith C.A., Lazarus R.S. Emotion and Adaptation. In: Pervin L.A. ed. Handbook of Personality: Theory and Research, Guilford, New York, 1990. P. 609-637.

8. Li X., Chen M. Multimodal sentiment analysis with multi-perspective fusion network focusing on sense attentive language // C hinese Computational Linguistics: 19th China National Conference, CCL 2020, Hainan, China, October 30 – November 1, 2020, Proceedings 19. Springer International Publishing, 2020. P. 359-373.

9. Wen L., Li X., Gao L. A transfer convolutional neural network for fault diagnosis based on ResNet-50 // Neural Computing and Applications. 2020. Vol. 32. № 10. P. 6111-6124.

10. ImageNet Large Scale Visual Recognition Challenge 2013 (ILSVRC2013) // ImageNet. [Электронный ресурс]. URL: https://image-net.org/challenges/LSVRC/2013/ (дата обращения: 27.04.2024).

11. Yu Y., Si X., Hu C., Zhang J. A review of recurrent neural networks: LSTM cells and network architectures // Neural computation. 2019. Т. 31, № 7. С. 1235-1270.

12. Staudemeyer R. C., Morris E. R. Understanding LSTM--a tutorial into long short-term memory recurrent neural networks // arXiv preprint arXiv:1909.09586. 2019. [Электронный ресурс]. URL: https://arxiv.org/pdf/1909.09586 (дата обращения: 03.04.2024).

13. Шамарина Е.А., Гусева А.И., Киреев В.С. Сентимент-анализ на основе нейросетей-трансформеров // Нейроинформатика-2023: сборник научных трудов XXV Международной научно-технической конференции. М., 2023. С. 292-301.

14. Devlin J., Chang M.W., Lee K., Toutanova K. Pre-training of deep bidirectional transformers for language understanding // arXiv preprint arXiv:1810.04805. 2018. [Электронный ресурс]. URL: https://arxiv.org/pdf/1810.04805 (дата обращения: 21.04.2024).

15. Martin O., Kotsia I., Macq B., Pitas I. The eNTERFACE 05 Audio-Visual Emotion Database, Proceedings of the First IEEE Workshop on Multimedia Database Management, Atlanta, April 2006. [Электронный ресурс]. URL: https://typeset.io/pdf/the-enterface05-audio-visual-emotion-database-57yw98sre5.pdf (дата обращения: 06.04.2024). DOI: 10.1109/ICDEW.2006.145.