УДК: 004.8:72:004.9

АНАЛИЗ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ И КОММУНИКАТИВНЫХ СПОСОБНОСТЕЙ В СИСТЕМАХ ЧЕЛОВЕК–МАШИНА С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

Гусев М.Д. – студент ИнПИТ
Шонин Д.В. – ассистент кафедры ПИТ
Кулагин И.М. – ассистент кафедры ПИТ
Саратовский Государственный Технический Университет
имени Ю.А. Гагарина
Россия, г. Саратов

Человек в любой момент времени находится в определённом эмоциональном состоянии. Это состояние влияет на внимание, скорость реакции, принятие решений и качество взаимодействия, то есть на так называемый человеческий фактор. В технических и организационных системах, где роль человека велика, учет человеческого фактора необходим для повышения надёжности и эффективности работы [1].

Минимизация негативного влияния человеческого фактора невозможна без его анализа и прогнозирования. Знание текущего состояния и его динамики позволяет выявлять аномалии, связанные, например, с повышенным напряжением, усталостью или агрессией, и заранее снижать риски.

Любая социально-техническая система состоит из множества взаимодействующих элементов. Люди в такой системе связаны между собой коммуникацией, и качество этой коммуникации напрямую влияет на результат работы. Коммуникативные навыки (soft skills) включают структурированность речи, ясность формулировок, умение слушать и отвечать, кооперативность и соблюдение норм общения.

Коммуникативные способности отдельного человека влияют не только на результат совместной деятельности, но и на эмоциональное состояние коллектива, усиливая или, наоборот, сглаживая проявления человеческого фактора. Отсюда следует необходимость их оценки и последующего развития в составе общей системы управления.

Определение эмоционального состояния и коммуникативных навыков позволяет прогнозировать риски, связанные с человеческим фактором, и повышать эффективность систем человек–машина и коллективной работы.

Современные нейронные сети способны решать широкий круг задач, однако их применение оправдано не всегда. Для части задач (решение аналитических формул, форматирование текста по жёстким правилам) проще и точнее использовать алгоритмические методы. Использование нейросетей целесообразно там, где:

1. **Невозможно задать чёткий алгоритм решения.**

2. **Существует достаточно примеров** решения задачи или ясное понимание соответствия входов и выходов.

3. **Между элементами задачи есть сложные связи**, доступные человеческому восприятию, но трудные для формального описания.

Анализ эмоционального состояния и коммуникативных навыков как раз относится к такому классу задач. Для него имеются большие массивы аудио-, видео- и текстовых данных, но отсутствуют простые правила, позволяющие однозначно вывести эмоцию или коммуникативное качество из отдельных признаков. Современные обзоры по мультимодальному распознаванию эмоций подтверждают, что именно нейросетевые методы позволяют эффективно моделировать такие зависимости [2–4].

При живом общении человек использует несколько каналов передачи информации:

- **интонация** (аудио);
- **жесты и поза**;
- **мимика**;
- **текст/содержание речи**.

Слова несут смысл, но интонация, мимика и жесты придают высказыванию эмоциональную окраску и дополняют информацию о коммуникативном стиле. Для адекватного анализа необходимо учитывать все каналы одновременно или максимально возможное их количество.

Для каждой модальности применимы специализированные нейросетевые модели:

- **Интонация.** Аудиосигнал преобразуется в набор акустических признаков или спектрограмм, по которым сеть оценивает признаки эмоций (уверенность, возбуждение, раздражение и др.). Предварительная обработка (шумоподавление, нормализация) повышает точность [2].
- **Жесты.** По видеоизображению оцениваются положение и движения частей тела. Резкие и хаотичные движения могут указывать на нервозность, закрытая поза — на неуверенность или нежелание к взаимодействию.
- **Мимика.** Выражение лица – один из базовых носителей эмоциональной информации. Сети компьютерного зрения по последовательности кадров определяют вероятности базовых эмоций или координаты в непрерывном эмоциональном пространстве [2, 4].
- **Текст.** Распознанная речь используется для оценки эмоциональной окраски и коммуникативных характеристик: логичности, ясности, агрессивности, кооперативности. Языковые модели позволяют учитывать контекст и структуру диалога [3, 5].

Каждый из способов по отдельности даёт неполную и шумную оценку. Нейронные сети не могут гарантировать 100% точности, поэтому ключевым является **совместный анализ** нескольких источников информации.

Для объединения результатов по отдельным каналам вводится модуль **мультимодального слияния** (мультиплексор). На его вход поступают:

- оценки эмоций и коммуникативных показателей по каждому каналу;

● информация о качестве данных (наличие шума, пропуски, отсутствие модальности).

Основные функции мультиплексора:

1. **Сопоставление данных** во времени и по контексту, выявление противоречий между каналами (например, нейтральные слова при агрессивной интонации).

2. **Объединение данных** с учётом весов модальностей, зависящих от их информативности и качества в конкретной ситуации (меньший вес жестов по сравнению с мимикой и текстом и т.п.).

3. **Структурирование результатов** в виде интегральных показателей эмоционального состояния и коммуникативных навыков, удобных для восприятия человеком.

Современные обзоры по мультимодальной обработке эмоций показывают, что продуманные стратегии слияния модальностей позволяют существенно повысить устойчивость и точность по сравнению с использованием одного канала [4, 5].

При практическом применении необходимо учитывать этические и правовые ограничения: работа с аудио- и видеоданными требует защиты персональной и биометрической информации, информированного согласия пользователей и запрета на использование системы как единственного критерия принятия решений о человеке [1].

Определение эмоционального состояния и коммуникативных навыков человека является важной задачей в контексте систем человек–машина и управления человеческим фактором. Невозможность задать жёсткий алгоритм решения и наличие большого числа примеров делают эту задачу естественной областью применения нейронных сетей.

Предложенный подход опирается на анализ четырёх каналов — интонации, мимики, жестов и текста — и использование модуля мультимодального слияния (мультиплексора), который формирует итоговые оценки. Модульная архитектура позволяет адаптировать систему к различным предметным областям и условиям (наличие/отсутствие отдельных модальностей), а также интегрировать её в существующие информационные решения.

Дальнейшие исследования могут быть направлены на разработку прототипов для конкретных доменов (управление персоналом, образование, операторские рабочие места), сбор специализированных мультимодальных датасетов и более глубокий анализ этических аспектов применения подобных систем.

Список литературы:

1. Vinciarelli A., Pantic M., Bourlard H. Social signal processing: Survey of an emerging domain // Image and Vision Computing. – 2009. – Vol. 27, No. 12. – P. 1743–1759.

2. Wu Y., et al. A Comprehensive Review of Multimodal Emotion Recognition // Sensors. – 2025. – Article 12292624.

3. Ahmed N., et al. A systematic survey on multimodal emotion recognition using learning algorithms // Intelligent Systems with Applications. – 2023. – Vol. 17.

4. Lian H., et al. A Survey of Deep Learning-Based Multimodal Emotion Recognition // Entropy. – 2023. – Vol. 25, No. 10. – Article 1440.

5. A Survey of Multi-modal Emotion Recognition Based on Fusion of Facial Expression, Voice Intonation and Text Analysis // HSET. – 2024. – No. 12.