

**ФОРМИРОВАНИЕ НАБОРА ИНДИВИДУАЛЬНЫХ
ЭМОЦИОНАЛЬНЫХ ПРИЗНАКОВ С
ИСПОЛЬЗОВАНИЕМ БИНАРНОЙ КЛАССИФИКАЦИИ
ПОЛЬЗОВАТЕЛЬСКИХ ТЕКСТОВ**

***Савенков Павел Анатольевич, Волошко Анна
Геннадьевна, Крюков Олег Сергеевич***

доцент, доцент, аспирант

Тульский государственный университет, Тула, Россия

E-mail: pavel@savenkov.com, atroshina@mail.ru, ol_kryukov97@mail.ru

Научный руководитель — Ивутин Алексей Николаевич

В условиях стремительного развития цифровых технологий и увеличения объемов пользовательских данных, анализ текстовой информации приобретает особую значимость. Выявление эмоциональных и поведенческих паттернов на основе текстовых данных становится востребованным инструментом в таких областях, как информационная безопасность, управление персоналом и поведенческий анализ [1]. Данный вид анализа способен минимизировать риски, связанные с инсайдерской активностью сотрудников или несанкционированным доступом к корпоративной информации [2].

В условиях активного использования мобильных устройств в компаниях, возникает потребность в инструментах, позволяющих оперативно оценивать эмоциональную окраску вводимых на них сообщений с целью выявления поведенческие паттернов. Анализ коротких текстов, вводимых пользователями на мобильных устройствах, имеет ряд особенностей, которые требуют применения новых методик обработки данных. Основная цель исследования – разработка метода, позволяющего классифицировать эмоциональную окраску сообщений и формировать на ее основе динамические профили пользователей. В настоящее время, существующие подходы к анализу текстов, в основном, делятся на лексико-ориентированные и методы машинного обучения. Лексико-ориентированные методы (SentiWordNet, AFINN) основаны на использовании заранее подготовленных словарей, содержащих эмоционально окрашенные слова. Однако эти методы оказываются малоэффективными, так как не учитывают контекст, что является важным фактором в коротких сообщениях. Например, выражение "не счастлив" может быть ошибочно интерпретировано как положительное. Методы машинного обучения позволяют учитывать контекст и выявлять сложные зависимости в данных. В последние годы особую популярность при-

обрели рекуррентные нейронные сети (RNN) и их улучшенные версии, а именно сети LSTM и GRU. Для оценки эффективности различных архитектур нейронных сетей, предварительно обученных на наиболее подходящем наборе данных «RuTweetCorp» было проведено экспериментальное исследование на основе анализа коротких текстов. В ходе экспериментов была протестирована модель с слоем SimpleRNN, которая достигла 79% точности на тестовых данных. Данная нейронная сеть способна учитывать последовательность слов, однако качество ее работы не находится на высоком уровне. Модель со слоем LSTM обеспечила 84% точности и показала, что может лучше учитывать контекст. Однако наилучший результат продемонстрировала архитектура GRU, которая достигла 87% точности. В сравнении с LSTM, GRU имеет менее сложную структуру, что ускоряет процесс обучения, снижает вероятность переобучения и при этом сохраняет высокое качество предсказаний. Эти характеристики делают GRU наиболее предпочтительной для анализа коротких текстов, именно для формирования набора индивидуальных эмоциональных признаков, обеспечивая баланс между точностью и затратами вычислительных ресурсов.

Анализ эмоциональной окраски коротких текстов напрямую не позволяет оценить полную картину активности сотрудников. В связи с этим необходимо собрать набор индивидуальных признаков, которые будут отражать характерные поведенческие особенности. Этот набор признаков в дальнейшем служит основой для визуализации, что позволяет оператору превентивно принимать решения, основываясь на предоставленных данных в виде массивов изменений поведения и графических представлений. На рис. 1 изображен интерфейс оператора собственной системы мониторинга «DeepViewer», в которую был интегрирован разработанный метод. На графике представлены изменения поведения сотрудника (положительные/отрицательные классы).

Основными этапами формирования наборов индивидуальных эмоциональных признаков с использованием бинарной классификации пользовательских текстов является: сбор текстов в течение заданного временного интервала, обработка текстов с использованием предобученной нейросети на базе GRU, классификация текстов на позитивные и негативные; формирование вектора признаков на основе количества обнаруженных эмоциональных характеристик, визуализация динамики изменений эмоционального состояния пользователя. Данный метод позволяет не только фиксировать текущее эмо-

циональное состояние, но и отслеживать его изменения во времени. В ходе работы был разработан метод формирования набора индивидуальных эмоциональных признаков на основе анализа коротких пользовательских текстов и интегрирован в собственную систему мониторинга сотрудников. Проведенное исследование продемонстрировало, что рекуррентные нейросети, в частности GRU, обеспечивают наилучшие результаты при обработке таких данных. Применение предложенного метода особенно актуально в системах, где важно учитывать эмоциональное состояние пользователей. В дальнейшем планируется исследование гибридных архитектур, объединяющих преимущества трансформеров и рекуррентных нейросетей, что возможно позволит дополнительно повысить эффективность анализа.

Иллюстрации

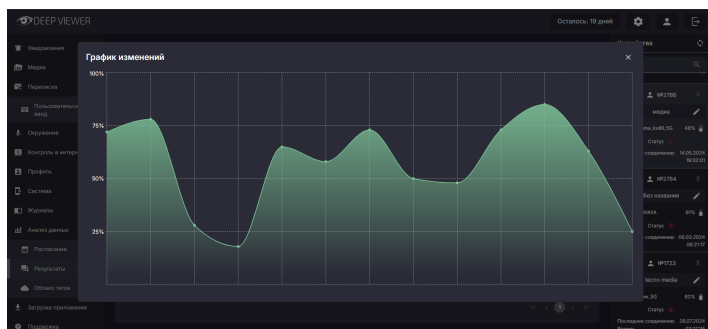


Рис. 1. Интерфейс оператора собственной системы мониторинга «DeepViewer» с графиком изменений поведения сотрудника

Литература

1. Davenport T. H. et al. Artificial intelligence for the real world // Harvard business review, 2018, T. 96, № 1. P. 217–226.
2. Park K. Kim H. J. Kim J. M. The effect of mobile device usage on creating text reviews // Asia Pacific Journal of Marketing and Logistics, 2023, T. 35, № 4. P. 1011–1029.