

МЕТОД МНОГОЗАДАЧНОГО РАСПОЗНАВАНИЯ ПСИХОФИЗИОЛОГИЧЕСКИХ СОСТОЯНИЙ ЧЕЛОВЕКА ПО РЕЧЕВОМУ СИГНАЛУ

Абдулкадиров Т.Н.¹

Научный руководитель – канд. техн. наук Рюмин Д.А.¹

¹Санкт-Петербургский Федеральный исследовательский центр Российской академии наук
timur.abdulkadirov2@mail.ru

Введение

Задача распознавания психофизиологических состояний человека является одной из ключевых для систем человеко-машинного взаимодействия. Несмотря на высокую эффективность мультимодальных методов, решения, основанные на речевом сигнале, являются наиболее практичными, поскольку речевые данные содержат большое количество информации и при этом хорошо собираются в естественных сценариях взаимодействия.

Большинство научных исследований сосредоточены на одной задаче, что требует машинного обучения отдельных нейросетевых моделей под каждую задачу (например, распознавание эмоций [1], оценка персональных качеств [2], распознавание сложных поведенческих состояний [3]). В качестве оптимального решения выступает многозадачное машинное обучение, позволяющее обучить единую нейросетевую модель сразу на нескольких задачах.

Основная часть

В работе предлагается метод многозадачного распознавания психофизиологических состояний человека по речевому сигналу. На первом этапе входной речевой сигнал преобразуется с помощью предобученного энкодера для получения устойчивого акустического представления фиксированной размерности.

Далее эмбединги подаются в общий для всех задач энкодер, формирующий единое представление, содержащее информацию, полезную для всех рассматриваемых задач. В качестве такого энкодера рассматриваются Трансформер и Mamba, а также вариация трансформера с линейным вниманием (вычислительно эффективным механизмом внимания) [4], что позволяет исследовать влияние различных механизмов моделирования последовательностей на качество итоговой нейросетевой модели, после чего для каждой задачи используется отдельная выходная голова, реализующая отображение общего представления в пространство целевых меток соответствующей задачи. Такое разделение архитектуры позволяет совместно использовать общее представление признаков при сохранении специфики отдельных задач.

Дополнительно в работе анализируются стратегии балансировки и оптимизации многозадачного машинного обучения (для решения проблем наивного многозадачного машинного обучения: негативного переноса, дисбаланса оптимизации, конфликта градиентов и др.), включая взвешивание задач и согласование градиентов для стабилизации машинного обучения, переноса знаний между смежными задачами, повышения обобщающей способности и улучшения качества итоговой нейросетевой модели [5-7]. Сравнение различных стратегий многозадачной оптимизации позволяет оценить их влияние на устойчивость машинного обучения и качество предсказаний по отдельным задачам.

Выводы

Результаты показывают существенное влияние архитектуры нейросетевой модели и стратегии многозадачного обучения на получаемое качество. Вычислительно простые нейросетевые модели демонстрируют сопоставимые результаты, а применение методов

балансировки позволяет снизить негативный перенос на более сложных задачах и стабилизировать процесс машинного обучения.

Литература

1. Zadeh A.B., Liang P.P., Poria S., Cambria E., Morency L.P. Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph // Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL). 2018. P. 2236-2246. DOI: 10.18653/v1/P18-1208.
2. Escalante H.J., Kaya H., Salah A.A., Escalera S., Güçlütürk Y., Güçlü U., Baró X., Guyon I., Jacques Junior J.C.S., Madadi M., Ayache S., Viegas E., Gürpınar F., Wicaksana A.S., Liem C.C.S., van Gerven M.A.J., van Lier R. Modeling, Recognizing, and Explaining Apparent Personality from Videos // IEEE Transactions on Affective Computing. 2020. Vol. 13, No. 2. P. 894-911. DOI: 10.1109/TAFFC.2020.2973984.
3. González-González M., Belharbi S., Zeeshan M.O., Sharafi M., Aslam M.H., Pedersoli M., Koerich A.L., Bacon S.L., Granger E. BAH Dataset for Ambivalence/Hesitancy Recognition in Videos for Behavioural Change // International Conference on Learning Representations (ICLR). 2026. P. 1-46.
4. Lu J., Han X., Sun Y., Pati V., Kim Y., Somani S., Yang S. ZeroS: Zero-Sum Linear Attention for Efficient Transformers // Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS). 2025. P. 1-27.
5. Qin X., Wang X., Liao N., Yan J. NTKMTL: Mitigating Task Imbalance in Multi-Task Learning from Neural Tangent Kernel Perspective // Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS). 2025. P. 1-29.
6. Jeong W., Yoon K.J. Selective Task Group Updates for Multi-Task Optimization // International Conference on Learning Representations (ICLR). 2025. P. 1-34.
7. Ban H., Ji K. Fair Resource Allocation in Multi-Task Learning // Proceedings of the 41st International Conference on Machine Learning (ICML 2024). PMLR, 2024. arXiv:2402.15638.