

МУЛЬТИМОДАЛЬНОЕ РАСПОЗНАВАНИЕ ЭМОЦИЙ В ДИАЛОГАХ С ИСПОЛЬЗОВАНИЕМ ГРАФОВЫХ НЕЙРОННЫХ СЕТЕЙ

Смирнов Е. А.,

Научный руководитель – канд. техн. наук, доцент Коротеева О. В.

Университет ИТМО

Введение

В последние годы задачи распознавания эмоций в диалогах привлекают все больше внимания в области искусственного интеллекта. Технологии распознавания эмоций применяются в образовательной и корпоративной сфере, в маркетинге и рекламе для анализа клиентского опыта и в игровой индустрии для адаптивного повествования.

Современные методы классификации эмоций, в том числе мультимодальные подходы, часто анализируют реплики изолированно [1, 2], не учитывая в должной мере диалоговый контекст, смену дикторов и темпоральные характеристики взаимодействия. Перспективным решением являются графовые нейронные сети, способные моделировать взаимодействия между участниками.

Основная часть

Предлагается подход к распознаванию эмоций в диалогах на основе мультимодальных графовых нейронных сетей. Диалог моделируется в виде динамического графа, в узлах которого находятся реплики, а ребра отражают не только семантические и синтаксические связи, но и учитывают хетизации (временные паузы) и контекстные связи между репликами. Для обучения модели используются текстовая и аудио модальности. Текстовые признаки извлекаются с помощью предобученных языковых моделей, а акустические на основе стандартных дескрипторов (мел-кепстральные коэффициенты, высота тона и др.). Для объединения модальностей применяются стратегии раннего и позднего слияния.

В ходе экспериментального исследования на стандартном датасете MELD [3] было показано, что графовая нейронная сеть с учетом контекстных и временных признаков диалога повышает точность классификации эмоций по сравнению с базовыми методами.

Выводы

Предложенный подход демонстрирует эффективность представления диалога в виде графовой структуры для задачи распознавания эмоций. Учет глобального контекста диалога и хетизации позволяет более точно моделировать эмоциональную динамику общения и повышает качество классификации по сравнению с подходами, не учитывающими структуру диалога. Полученные результаты могут быть полезны для дальнейших исследований в области речевой аналитики и построения диалоговых систем.

Литература

1. Sahil S., Sangwan A., Mishra A. COGMEN: Contextualized GNN based Multimodal Emotion Recognition. – 2022. [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2205.02455>, свободный. – Яз. англ. (дата обращения: 01.02.2025).
2. Shen W., Gao X., He X. Graph Neural Networks for Dialogue Emotion Recognition. – 2022. [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1908.11540>, свободный. – Яз. англ. (дата обращения: 21.01.2025).

3. Poria S., Hazarika D., Majumder N., Naik G., Cambria E., Mihalcea R. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations. – 2018. [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1810.02508>, свободный. – Яз. англ. (дата обращения: 01.02.2025).