

## РАЗРАБОТКА ИНСТРУМЕНТА АНАЛИЗА ОБРАТНОЙ СВЯЗИ МЕТОДОМ ДОМЕННОЙ АДАПТАЦИИ

Дронова А. С.<sup>1</sup>

Научный руководитель – к.т.н., доцент, зам. директора ВШЦК Романов А. А.<sup>1</sup>

<sup>1</sup>Университет ИТМО

dr.aleksandra1717@yandex.ru

### Введение

В условиях насыщения рынка онлайн-образования (EdTech) и снижения темпов его роста образовательные платформы стремятся фокусироваться на улучшении пользовательского опыта. Источником идей доработок продукта становится текущая база клиентов, а главным инструментом оценки качества продукта — NPS-опросы. Однако ручной анализ десятков тысяч отзывов требует колоссальных трудозатрат и не позволяет оперативно реагировать на «боли» клиентов. В связи с этим возникает научная и прикладная проблема автоматизации аспектно-сентиментного анализа (ABSA) — извлечения конкретных тем из текста и определения их тональности.

Анализ зарубежного и отечественного опыта показывает, что классические дискриминативные модели (на базе архитектур BiLSTM или BERT) обладают низкой способностью к обобщению на новые темы и плохо справляются с неявными аспектами, которые часто встречаются в естественном языке. Современным стандартом становится использование генеративных моделей архитектуры Encoder-Decoder (в частности, семейства G5), где задача извлечения аспектов формулируется как генерация структурированного текста. Тем не менее, применение базовых генеративных моделей к текстам узкоспециализированных доменов сталкивается с проблемой «доменного сдвига», что значительно влияет на качество обработки обратной связи.

### Основная часть

Данная работа посвящена разработке инструмента для автоматического извлечения полярных групп аспектов (позитив и негатив) из неструктурированных отзывов пользователей образовательной платформы. В качестве базовой архитектуры выбрана модель gpt5-base. Для преодоления проблемы ограниченных вычислительных ресурсов и семантического разрыва предложен комплексный пайплайн обучения, состоящий из двух ключевых этапов.

Первым этапом предложено использование метода доменной адаптации (Domain-Adaptive Pre-Training, DAPT). На неразмеченном корпусе из примерно 20 000 отзывов пользователей платформа модель обучалась задаче восстановления текста (Denoising): 15% слов заменялись на токен-маску, которую модель должна была восстановить из контекста. Это позволило сформировать устойчивые семантические связи для специфических терминов домена онлайн-образования (например, «куратор», «вебинар», «перевернутый класс», «бесшовное прикрепление»).

Второй этап подразумевал контролируемое дообучение (Supervised Fine-Tuning, SFT) на сбалансированном датасете из 3 000 размеченных отзывов. Для устранения естественного дисбаланса классов (превалирование позитивных отзывов над негативными) применялся метод обучения на контрастных примерах (Hard Negative Mining). Для оптимизации потребления видеопамяти использовался метод низкоранговой адаптации (LoRA). Область адаптации охватывала как проекционные матрицы внимания, так и полносвязные блоки, что затронуло 8% от общего числа параметров модели, но обеспечило высокую точность без эффекта катастрофического

забывания. Инференс модели реализован с применением метода лучевого поиска, гарантирующего генерацию связных и грамматически корректных конструкций.

### Выводы

В результате исследования разработан и обучен прикладной инструмент, способный структурировать сырой текст отзыва в списки позитивных и негативных тематических аспектов. Сравнительный анализ показал высокое превосходство предложенного подхода (DAPT + SFT) над базовым дообучением (только SFT). Внедрение этапа доменной адаптации позволило повысить интегральную метрику семантической близости предсказанных аспектов (Semantic F1) на 14%, а метрику лексической точности (BLEU) — на 62%. Кроме того, адаптированная модель на 45% чаще использует специфическую терминологию предметной области, избегая тривиальных обобщений.

Особую практическую ценность представляет рост метрики полноты выявления негативных аспектов более чем на 8%. Распределение разницы Semantic F1 показало, что разработанный подход успешно справляется с краевыми случаями, сложной доменной лексикой и неявным недовольством пользователей. Результаты исследования в настоящее время внедряются в качестве аналитического модуля для регулярного автоматизированного ревью продуктов.

### Литература

1. Gururangan, S. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks [Электронный ресурс] / S. Gururangan, A. Marasović, S. Swayamdipta [et al.] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – 2020. – P. 8342–8360. – URL: <https://aclanthology.org/2020.acl-main.740> (дата обращения: 06.12.2025).
2. He, Z. A Survey on Recent Advances in Sequence Labeling from Deep Learning Models [Электронный ресурс] / Z. He, Z. Wang, W. Wei [et al.] // arXiv.org. – 2020. – URL: <https://arxiv.org/pdf/2011.06727> (дата обращения: 06.12.2025).
3. Hu, E. J. LoRA: Low-Rank Adaptation of Large Language Models [Электронный ресурс] / E. J. Hu, Y. Shen, P. Wallis [et al.] // arXiv.org. – 2021. – URL: <https://arxiv.org/abs/2106.09685> (дата обращения: 06.12.2025).
4. Raffel, C. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer [Электронный ресурс] / C. Raffel, N. Shazeer, A. Roberts [et al.] // Journal of Machine Learning Research. – 2020. – Vol. 21, № 140. – P. 1–67. – URL: <http://jmlr.org/papers/v21/20-074.html> (дата обращения: 06.12.2025).
5. Yan H. [et al.]. A Unified Generative Framework for Aspect-Based Sentiment Analysis [Электронный ресурс] // Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics. – 2021. – P. 2416–2429. – URL: <https://aclanthology.org/2021.acl-long.188> (дата обращения: 08.12.2025)
6. Zhang W. [et al.]. Aspect Sentiment Quad Prediction as Paraphrase Generation [Электронный ресурс] // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. – 2021. – P. 4803–4813. – URL: <https://aclanthology.org/2021.emnlp-main.381> (дата обращения: 08.12.2025).