

МЕТОДЫ ГЕНЕРАЦИИ И ВЕКТОРИЗАЦИИ ПЕРСОНАЛИЗИРОВАННЫХ ОБРАЗОВАТЕЛЬНЫХ ТРАЕКТОРИЙ С ИСПОЛЬЗОВАНИЕМ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Новосельцев А. А.¹, Скобликов К. А.¹, Короткова А. А.¹

Научный руководитель – ассистент Терещенко В. В.¹

¹Университет ИТМО

andrew.nvst@yandex.ru

Введение

Современные программы обучения, чаще всего, предлагают унифицированный учебный план для всех студентов, независимо от их увлечений и интересов. Это приводит к тому, что часть аудитории вовсе не заинтересована в процессе обучения. Проблема персонализации образовательных траекторий остаётся актуальной: ручное составление индивидуальных планов трудозатратно и плохо масштабируется.

Появление больших языковых моделей (БЯМ) открыло возможность автоматической генерации учебных планов под конкретного студента. Однако применение этих моделей в образовательном контексте сопряжено с рядом нетривиальных вопросов: как сравнивать между собой траектории, полученные от разных моделей? Стандартные метрики наподобие BLEU и ROUGE здесь не подходят – они измеряют близость к эталонному тексту, тогда как в нашем случае заранее правильного ответа попросту нет. Дополнительно, появляется вероятность того, что сгенерированный план может не подходить под методические требования [1].

Основная часть

В работе реализованы два ключевых компонента: генерация образовательных траекторий с помощью БЯМ и их векторизация для дальнейшего сравнения и обработки.

Для формирования образовательных траекторий была сформирована достаточная сводка про студента. Полученные генерации проверялись на формальности на начальном этапе с помощью промпт-инженерии. Дополнительно были проведены эксперименты с различными моделями, в ходе которых были выявлены типовые ошибки: нарушение ограничений и несоблюдение формата ответа [2].

Модуль векторизации был построен по принципу единого интерфейса: независимо от типа входных данных, все векторизаторы предоставляли одинаковые методы fit, transform и save/load. Для текстовых описаний курсов было реализовано два подхода - TF-IDF с лемматизацией и фильтрацией стоп-слов, а также Sentence-Transformer. Категориальные признаки обрабатывались через OneHotVectorizer либо TargetEncoderVectorizer в зависимости от мощности категории. Числовые показатели – трудоёмкость, оценки, процент прохождения – нормировались NumericVectorizer с несколькими стратегиями масштабирования и заполнением пропусков [3].

Выводы

Реализованная система позволяет автоматически генерировать несколько вариантов образовательной траектории под описание конкретного студента и приводить их к единому векторному пространству, пригодному для дальнейшего анализа.

Среди направлений дальнейшей работы – сравнение качества траекторий от разных БЯМ с привлечением экспертной оценки, разработка педагогически обоснованных метрик и пилотная апробация системы на студентах технических специальностей в связке с реальной LMS.

Литература

1. Wang T. и др. LLM-powered Multi-agent Framework for Goal-oriented Learning in Intelligent Tutoring System // Companion Proceedings of the ACM on Web Conference 2025. Sydney NSW Australia: ACM, 2025. С. 510–519.
2. Zhang X. и др. EduPlanner: LLM-Based Multi-Agent Systems for Customized and Intelligent Instructional Design: arXiv:2504.05370. arXiv, 2025.
3. Gao X. и др. GoAI: Enhancing AI Students' Learning Paths and Idea Generation via Graph of AI Ideas: arXiv:2503.08549. arXiv, 2025.