

## СОВЕРШЕНСТВОВАНИЕ АЛГОРИТМА ОСТАТОЧНОЙ КВАНТИЗАЦИИ ДЛЯ ПОЛУЧЕНИЯ СЕМАНТИЧЕСКИХ ИДЕНТИФИКАТОРОВ

*Сорокин Олег Владиславович*

*Студент*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: ovs02@mail.ru*

*Научный руководитель — Ищенко Роман Валерьевич*

В рекомендательных системах в настоящее время наблюдается активный переход к генеративному подходу, в рамках которого модель учится порождать представления релевантных предметов из заданного каталога. В качестве таких представлений могут служить семантические идентификаторы — целочисленные последовательности, кодирующие признаковые описания предметов [2].

Для построения семантических идентификаторов часто используется алгоритм RQVAE, основанный на остаточной квантизации [2]. Известно, что учёт коллаборативного сигнала при построении семантических идентификаторов может существенно повышать качество рекомендаций. Один из распространённых подходов состоит во введении дополнительной компоненты в функцию потерь при обучении RQVAE, которая сближает представления предметов, связанных пользовательским поведением [1]. В настоящей работе предлагается ряд модификаций этого подхода.

Рассматривается вопрос о выборе коллаборативных пар для обучения: предложен критерий минимальной информативности пары, позволяющий отбирать только статистически значимые связи и тем самым снижать долю шума в обучающей выборке.

Показано, что добавление коллаборативной компоненты на практике может приводить к неэффективному использованию параметров модели-квантизатора. В частности, в ходе остаточной квантизации возникает дисбаланс выбора векторов: одни и те же элементы используются существенно чаще остальных. Кроме того, часть векторов может не использоваться. Для уменьшения числа неиспользуемых элементов предложена процедура переинициализации редко используемых центров, приводящая к более стабильной динамике обучения. Для уменьшения дисбаланса разработана специальная добавка к функции потерь, основанная на идее балансировки смесей экспертов в современных языковых моделях.

Проведены эксперименты на открытых наборах данных. Показано, что предложенные модификации приводят к улучшению качества рекомендаций на тестовой выборке.

### Литература

1. He R., Heldt L., Hong L. et al. PLUM: Adapting Pre-trained Language Models for Industrial-scale Generative Recommendations // arXiv. 2025. URL: <https://arxiv.org/abs/2510.07784>
2. Rajput S., Mehta N., Singh A. et al. Recommender Systems with Generative Retrieval // In: Advances in Neural Information Processing Systems. 2023.