

УСЛОВНОЕ СЕМПЛИРОВАНИЕ НЕГАТИВНЫХ ПРИМЕРОВ ДЛЯ ОБУЧЕНИЯ МОДЕЛЕЙ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Замятин Е. И.¹

Научный руководитель – канд. техн. наук, доцент Муравьев С. Б.¹

¹Университет ИТМО
ezamyatin@itmo.ru

Введение

Распространенный способ обучения моделей рекомендательных систем заключается в сведении к задаче многоклассовой классификации, где каждый объект рекомендаций представляет свой класс. Современные информационные системы промышленного масштаба зачастую содержат сотни миллионов объектов рекомендаций, что делает невозможным обучение с использованием стандартных техник. Распространенное решение в таких случаях подразумевает использования методов негативного семплирования. Вместо вычисления функции потерь по всем объектам рекомендаций, используется небольшая подвыборка. Один из наиболее эффективных способов такого семплирования называется «внутрипакетные негативы» [1]. Его суть в выборе позитивных примеров других пользователей из мини-выборки в качестве негативных для текущего пользователя. В данной работе я предлагаю расширение данной техники, которое позволяет проводить семплирование из распределений, обусловленных на контекст.

Основная часть

В данной работе рассматривается проблема получения внутрипакетных негативов с условием на некоторый контекст. Проблема применения стандартной техники заключается в том, что при большой кардинальности контекстуальной переменной, разнообразие негативных примеров сильно снижается, за счет ограниченного размера мини-выборки. Чтобы решить эту проблему, я предлагаю концепцию потокового внутрипакетного семплирования. Его суть заключается в том, что вместо использования позитивов других пользователей мини-выборки, используется специальная структура, которая аккумулирует позитивы с предыдущих мини-выборок. За счет того, что на каждое значение контекстуальной переменной можно хранить собственную структуру, количество негативных примеров для каждого контекста фактически не ограничено.

Для проведения экспериментов с использованием разработанного алгоритма была взята открытая база данных рекомендательных систем MIND [2]. База данных представляет из себя набор взаимодействий пользователей с новостями. В качестве контекста негативного семплирования использовалась дата взаимодействия. Использование такого контекста позволило отделить эффект сезонности от релевантности, за счет чего результирующая метрика Recall@10 была существенно улучшена.

Выводы

В данной работе был представлен метод внутрипакетного негативного семплирования с условием на контекст. Применение метода на задаче рекомендаций новостей продемонстрировало его высокую эффективность и точность предсказаний результирующей модели.

Литература

1. Yi X. et al. Sampling-bias-corrected neural modeling for large corpus item recommendations //Proceedings of the 13th ACM conference on recommender systems. – 2019. – С. 269-277.
2. Wu F. et al. Mind: A large-scale dataset for news recommendation //Proceedings of the 58th annual meeting of the association for computational linguistics. – 2020. – С. 3597-3606.