

ИСПОЛЬЗОВАНИЕ СЛУЧАЙНЫХ БЛУЖДАНИЙ ПО ГРАФУ ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА РЕКОМЕНДАЦИЙ

Коваленко Павел Антонович

Студент

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

E-mail: pashakovalenko@yandex.ru

Научный руководитель — Дьяконов Александр Геннадьевич

В данной работе предлагается метод повышения точности рекомендательных систем. Идея метода заключается в увеличении объема обучающей выборки за счет добавления искусственных положительных примеров. Метод может быть успешно применен к различным задачам и алгоритмам ранжирования, что подтверждается проведенными экспериментами.

Рассматривается задача рекомендаций с неявным откликом — в обучающей выборке есть только положительные примеры. Обучающая выборка представляет собой набор пар (запрос, документ), где документы считаются релевантными запросу.

Введем двудольный неориентированный граф, в котором вершинам первой доли соответствуют запросы, вершинам второй доли — документы, запрос соединен ребром с документом, если документ релевантен запросу — пара (запрос, документ) встречается в обучающей выборке.

Введем случайное блуждание по графу следующим образом [5]. Блуждание начинается с некоторой вершины v_0 , на последующих шагах вершина v_i выбирается случайно из соседей вершины v_{i-1} . Вероятность перехода в вершину зависит от степени вершины. Случайное блуждание может быть фиксированной длины, либо после каждого шага есть некоторая вероятность останова.

Построим новую обучающую выборку следующим образом. В качестве начальной вершины выберем случайный запрос. Из него запустим случайное блуждание, наложив ограничение, чтобы оно закончилось в вершине, соответствующей документу. Полученную пару (первый запрос, последний документ) добавим в выборку. Описанным образом соберем достаточно большую новую выборку, которая и будет использована для обучения алгоритмов ранжирования.

Для подтверждения работоспособности предложенного метода был проведен ряд экспериментов на датасетах *MovieLens 1m* [1],

MovieLens 20m [1] и *Million Playlist Dataset* [4]. В качестве алгоритмов ранжирования были рассмотрены *Implicit Alternating Least Squares* [2] и *Bayesian Personalized Ranking* [3]. Метрика качества — *precision@10*. Точность для алгоритмов, обученных на исходных датасетах и на датасетах, расширенных при помощи случайных блужданий (*Random Walks*), представлена в таблице ниже. Для предложенного метода указан лучший результат среди возможных параметров метода.

Алгоритм	MovieLens 1m	MovieLens 20m	Million Playlist
ALS	0.1255	0.1174	0.1556
RW + ALS	0.1365	0.1308	0.1783
BPR	0.1345	0.1314	0.1842
RW + BPR	0.1942	0.1875	0.2196

Из таблицы видно, что для рассмотренных датасетов и алгоритмов предложенный метод позволяет значительно повысить точность ранжирования.

Литература

1. Harper F. M., Konstan J. A. The MovieLens Datasets: History and Context //ACM Transactions on Interactive Intelligent Systems (TiiS). – 2016. – Т. 5. – №. 4. – С. 19.
2. Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //Eighth IEEE International Conference on Data Mining. – IEEE, 2008. – С. 263-272.
3. Rendle S. et al. BPR: Bayesian Personalized Ranking from Implicit Feedback //Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. – AUAI Press, 2009. – С. 452-461.
4. Schedl M. et al. RecSys Challenge 2018: Automatic Playlist Continuation //Proceedings of the 18th International Society for Music Information Retrieval Conference. – 2017.
5. Yang J. H. et al. HOP-Rec: High-Order Proximity for Implicit Recommendation //Proceedings of the 12th ACM Conference on Recommender Systems. – ACM, 2018. – С. 140-144.