

УДК 004.62

Повышение точности прогнозов коллаборативной рекомендательной системы с помощью расчета евклидового расстояния и коэффициента Отиаи

Медведев Дмитрий Игоревич dmitriy.medvedev.work@yandex.ru

Кондратьева Светлана Дмитриевна sdkond@bmstu.ru

КФ МГТУ им. Н.Э. Баумана, Калуга, Россия

Рассмотрен комплексный метод расчета евклидового расстояния и коэффициента Отиаи в рекомендательных системах, основанных на коллаборативной фильтрации. Проанализированы методы определения схожести между оценками пользователей на основе данных мер, их совместное использование и влияние на конечные рекомендации. При этом уделяется внимание принципу работы алгоритма коллаборативной фильтрации. Цель работы — выявить метод, позволяющий улучшить качество рекомендаций с помощью коэффициента Отиаи.

Ключевые слова: рекомендательная система, коллаборативная фильтрация, Евклидово расстояние, коэффициент Отиаи

Введение. В современном мире существует множество сервисов, использующих рекомендательные системы. Одними из таких решений являются рекомендательные системы на основе коллаборативной фильтрации, обеспечивающие каждого пользователя персонализированными предложениями товаров и услуг, основываясь на его предпочтениях. Однако такие рекомендательные системы могут выдавать неточные оценки. Одной из причин данной проблемы является использование одного метода оценки. Применение комбинированного подхода с использованием коэффициента Отиаи позволит повысить точность прогнозов рекомендательной системы.

Принцип работы коллаборативной фильтрации. Коллаборативная фильтрация — это один из методов прогнозирования в рекомендательной системе, который применяется для предсказания неизвестных предпочтений пользователя на основе анализа известных предпочтений или оценок группы пользователей [1].

Идея коллаборативной фильтрации основывается на предположении, что пользователи всегда дают оценки по одному и тому же принципу. Также одним из основных подходов является суждение, что если у пользователей есть общие интересы, то их оценки будут ближе друг для друга, чем другие оценки.

Пусть имеется набор пользователей и объектов, которые они оценили. Это отражено в табл. 1. Задача заключается в том, чтобы «угадать», как, например, пользователь Д оценит объект З на основании оценок других пользователей. Для этого необходимо наиболее похожих на пользователя Д пользователей, оценивших данный объект. И наконец, на основании их оценок «предсказывается» оценка данного пользователя для данного объекта.

Таблица 1

Матрица оценок объектов пользователями

Пользователь	Объект 1	Объект 2	Объект 3	Объект 4	Объект 5
Пользователь А	5,0	–	3,0	–	4,0
Пользователь Б	–	–	1,0	–	2,0
Пользователь В	4,0	3,0	3,0	–	5,0
Пользователь Г	2,0	2,0	2,0	4,0	1,0
Пользователь Д	–	3,0	?	5,0	–
Пользователь Е	2,0	–	4,0	–	3,0

Среди коллаборативных алгоритмов выделяют основанные на пользователях (User-based) и на объектах (Item-based). В первом подходе сравниваются оценки пользователей, а во втором признаки объектов.

В качестве основных мер схожести выделяют: корреляцию Пирсона; Манхэттенское расстояние; коэффициент Танимото; евклидово расстояние; коэффициент Отиаи.

Рассмотрим в качестве мер схожести для рекомендательной системы расчет евклидового расстояния и коэффициента Отиаи.

Евклидово расстояние. Евклидово расстояние — это метрика, которая измеряет кратчайшее расстояние между двумя точками в евклидовом пространстве [2]. Формула Евклидового расстояния имеет следующий вид:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

В рекомендательных системах евклидово расстояние позволяет оценить, насколько оценки одного пользователя близки к оценкам другого пользователя, или насколько признак одного объекта похож на признак другого объекта. И чем меньше получится расстояние, тем ближе оценки друг к другу.

Однако при расчете Евклидового расстояния невозможно использовать по-разному признаки с разной шкалой измерений. Кроме того, при увеличении количества размерностей, эффективность метода падает. Данная проблема называют «Проклятием размерности» [3].

Если большинство оценок не заданы, то расчет евклидова расстояния на основе всего набора данных может быть искажен, так как результат будет зависеть от большого количества «нулевых» значений.

Коэффициент Отиаи. Коэффициент Отиаи измеряет угол между векторами предпочтений пользователей. Это позволяет учитывать отношения между предпочтениями, что делает его более устойчивым к проблемам, с ко-

торыми сталкиваются при расчете евклидового расстояния. Формула коэффициента Отиаи имеет следующий вид:

$$K = \frac{|A \cap B|}{\sqrt{|A| \times |B|}}.$$

Коэффициент Отиаи игнорирует разреженность данных, поскольку основана на углах между векторами, а не на абсолютных значениях элементов векторов [4].

Эксперимент. В качестве эксперимента созданы рекомендации для пользователя Д, рассчитанные с помощью евклидового расстояния и комбинированного подхода с применением коэффициента Отиаи. Датасетом является набор данных из таблицы 1. Несмотря на малый объем и разреженность датасета, его хватит для оценки работы методов. Результаты эксперимента отражены в табл. 2 и 3.

Таблица 2

Меры схожести пользователей

Метод расчета	Пользователь В	Пользователь Г	Сумма
Евклидово расстояние	0,1035	0,2317	0,3352
Комбинированный	0,0208	0,1918	0,2126

Таблица 3

Результаты работы рекомендательной системы

Метод расчета	Объект 1	Объект 2	Объект 3	Объект 4	Объект 5
Евклидово расстояние	2,62	3,0	3,31	5,0	2,236
Комбинированный	2,195	3,0	2,098	5,0	1,391

Эксперимент показал, что коэффициент Отиаи в комбинированном методе корректирует значения, учитывая схожесть рекомендаций пользователей. При этом, численное значение рекомендации в большей степени зависят от оценки наиболее похожих пользователей, за счет еще большего веса их оценок, чем при использовании только расчета евклидового расстояния.

Литература

[1] Белов Ю.С., Клюквин Р.В. Использование косинуса схожести в рекомендательных системах на основе коллаборативной фильтрации. *Наука, техника и образование*, 2017, № 2 (12), с. 131–136.

[2] *Евклидово расстояние*. URL: https://en.wikipedia.org/wiki/Euclidean_distance (дата обращения 19.10.2024).

- [3] Юре Л., Ананд Р., Джеффри Д.У. *Анализ больших наборов данных*. Москва, ДМК Пресс, 2016, 498 с.
- [4] Alammam J. The illustrated Word2vec. URL: <https://jalammar.github.io/illustratedword2vec/> (accessed October 22, 2024).

Improving the prediction accuracy of a collaborative recommender system by calculating the euclidean distance and the othiai coefficient

Dmitry I.M.

dmitriy.medvedev.work@yandex.ru

Kondratyeva S.D.

sdkond@bmstu.ru

BMSTU (KB), Kaluga, Russia

This paper considers a comprehensive method for calculating Euclidean distance and Othiai coefficient in recommendation systems based on collaborative filtering. The paper analyses the methods for determining the similarity between user ratings based on these measures, their joint use and their impact on the final recommendations. Attention is paid to the working principle of the collaborative filtering algorithm. The aim of the paper is to identify a method to improve the quality of recommendations using the Othiai coefficient.

Keywords: recommendation system, collaborative filtering, Euclidean distance, Ochiai coefficient, prediction improvement