International Journal of Information Retrieval Research
Volume 10 • Issue 1 • January-March 2020

Popularised Similarity Function for Effective Collaborative Filtering Recommendations

Abba Almu, Department of Mathematics, Computer Science Unit, Usmanu Danfodiyo University, Sokoto, Nigeria

Abubakar Roko, Usmanu Danfodiyo University, Sokoto, Nigeria

Aminu Mohammed, Department of Mathematics, Computer Science Unit, Usmanu Danfodiyo University, Sokoto, Nigeria

Ibrahim Saidu, Department of Information and Communication Technology, Faculty of Engineering and Environmental Desi, Sokoto, Nigeria

ABSTRACT

The existing similarity functions use the user-item rating matrix to process similar neighbours that can be used to predict ratings to the users. However, the functions highly penalise high popular items which lead to predicting items that may not be of interest to active users due to the punishment function employed. The functions also reduce the chances of selecting less popular items as similar neighbours due to the items with common ratings used. In this article, a popularised similarity function (pop_sim) is proposed to provide effective recommendations to users. The pop_sim function introduces a modified punishment function to minimise the penalty on high popular items. The function also employs a popularity constraint which uses ratings threshold to increase the chances of selecting less popular items as similar neighbours. The experimental studies indicate that the proposed pop_sim is effective in improving the accuracy of the rating prediction in terms of not only lowering the MAE but also the RMSE.

KEYWORDS

Modified Punishment Function, Popular Items, Popularised Similarity Algorithm, Popularity Constraint, Prediction

1. INTRODUCTION

The increasing amount of information on the web and the emergence of e-commerce have led to the problem of information overload. Because of this problem, it becomes difficult for users to search for items of their interest. Therefore, a recommender system is essential in order to identify items based on user's interest. The system is an information filtering system that recommends relevant items to users by analyzing the users explicitly mentioned preferences and interests (Ojokoh et al., 2013; Khusro, Ali & Ullah, 2016). It saves a lot of time and effort of users typically involve in issuing different queries about the items of interest, by simply prioritising and personalising large volume

DOI: 10.4018/IJIRR.2020010103

of information available at its disposal to find the unknown relevant items needed by the users. This prompted many research efforts on recommender systems (Goldberg et al., 1992; Burke, 2002; Linden, Smith & York, 2003; Miller et al., 2003; Pazzani & Billsus, 2007; Cacheda et al., 2011). Among these systems include Collaborative Filtering (CF) which is the most popular and successful system that provides recommendations to users because it recommends any type of items to users such as books, movies, news, music, web pages and so forth (Chen et al., 2011; Ping & Ming, 2012; Omisore, 2014).

The CF uses a similarity function to recommends items by considering users' ratings of an item to find the match of rating patterns of some items involving other users with similar interests (Montaner, Lopez & Rosa, 2003; Chen et al., 2011). The similarity function is the core component of the recommendation process because the accuracy defends on the type of similarity function used. Several similarity functions have been developed to predict the correct item rating based on the users preferences/ratings (Sarwar et al., 2001; Lee, Park & Park, 2007; Shen & Zhou, 2010; Weijie, Jing & Liang, 2012; Yang, Ali & Li, 2013; Zhao, Niu & Chen, 2013; Zhu et al., 2014; Latha & Nadarajan, 2015; Yang & Wang, 2016; Fan, Yu & Huang, 2018). Among these functions, the developed function in (Fan, Yu & Huang, 2018) uses a similarity function to punish popular items. However, the function may predict inaccurate interested popular items to the active users due to the penalty function employed by the similarity that highly penalises high popular items. In addition, it also leads to the return of few similar items neighbours because it utilises items with common ratings which decreases the tendency of selecting less popular items.

In this paper, a popularised similarity function (pop_sim) is proposed to provide effective recommendations to users. The pop_sim function introduces a modified punishment function to minimise the penalty on high popular items. The function also employs a popularity constraint which uses the items ratings threshold to increase the chances of choosing less popular items so as to obtain more similar items neighbours. Experimental studies conducted indicate that compared to the existing functions, the proposed pop_sim performs better in terms of reducing MAE and RMSE. Thus, improve the accuracy of the rating prediction.

The rest of the paper is organised as follows: Section 2 provides a brief review of related works on some existing similarity functions. The relevant preliminaries and concepts related to the CF algorithm are discussed in section 3. Section 4 presents our proposed similarity function in detail. The experimental evaluation and discussions on the results related to the accuracy of the proposed function are provided in section 5. Finally, the conclusion and future work are described in Section 6.

2. RELATED WORKS

In this section, a review on some existing similarity functions proposed to enhance the performance of recommender systems are presented. It focuses on highlighting operations, advantages and disadvantages of each function, which is as follows.

Sarwar et al. (2000) applied cosine similarity function in electronic commerce application to determine the similarity among two customers represented as two vectors containing users' ratings. The similarity function uses these vectors containing customers' ratings for the items to compute the cosine similarity which yields result between 0 and 1. It provides quality products recommendations to the target customers, but it fails to consider customers' preferences with different rating scales because it uses only two rating vectors.

Sarwar et al. (2001) presented a modified cosine similarity function to address the problem of users with various rating scales. The modified function uses explicit rating scales from 1 to 5 and then subtracts the users' average ratings for all rated items pair. The result turns to have less error while predicting the items to users and thus produce high quality recommendations than when the cosine similarity function is used. However, the function fails to capture the similarity from the implicit ratings of co-rated users or items.