Data-driven Study: Augmenting Prediction Accuracy of Recommendations in Social Learning Platforms

Soude Fazeli[a] Hendrik Drachsler[a] Peter Sloep[a]

[a] *Open Universiteit Nederland (OUNL), PO.Box 2960, The Netherlands*

Abstract

This study aims to develop a recommender system for a social learning platform to be provided by EU FP7 Open Discovery Space (ODS) project by taking into account social data of users to make recommendations. In this paper, we investigate which recommender algorithm can best fits social learning platforms like ODS platform. We conducted an experiment to test a set of different classical collaborative filtering algorithms on representative educational datasets similar to the future ODS dataset, as well as on the MovieLens dataset as a reference for studies on recommender systems. In addition to the classical collaborative filtering algorithms, we evaluated a graph-based recommender approach called T-index. We compare performance of the used algorithms in terms of F1 score. We also show how T-index approach can provide a balanced distribution of users' degree centrality.

1 The Goal

With the emergence of large amounts of data in various domains, recommender systems have become a practical approach to provide users with the most suitable information based on their interests and past behaviour. We apply recommender systems in the context of the FP7 Open Discovery Space[1] (ODS) project. The ODS contains large amounts of data in the field of education with a critical mass of approximately 1.550.000 eLearning resources from 75 content repositories, as well as 15 educational portals of regional, national or thematic coverage connected to it. Considering this huge amounts of data, we want to support ODS target users to find suitable content or people of their interest within ODS platform.

2 The Method

The first step to design a recommender system for ODS is to investigate what recommender algorithm best fits the ODS target users. To do so, we need to evaluate a set of recommender algorithms on ODS dataset including user social data e.g. rating, tagging, browsing, commenting, etc. Since we have no data yet from the ODS platform and its real users, we decided to conduct an offline empirical study for testing recommender algorithms on the datasets that are similar and related to the future ODS dataset. In the following sub sections, we describe the datasets and algorithms used for the offline data study.

2.1 Data

We selected the MACE[2] and OpenScout[3] datasets because the datasets contain social data of users such as ratings, tags, reviews, etc. on learning resources. So, their structure, content and target users are quite similar to the ODS datasets we aim to study. Running recommender algorithms on these datasets enables

---

[1] http://opendiscoveryspace.eu/index.php

[2] http://portal.mace-project.eu

[3] http://www.openscout.net/openscout-home

us to conduct an offline experiment in order to study the recommender algorithm to be customized for the ODS target users before going online with the actual users of the ODS. In addition to the above-mentioned educational datasets, we decided to use MovieLens[4] dataset (100K) as a reference dataset.

2.2 Algorithms

The educational datasets used in this study provide us with implicit preference values including browsing, tagging, commenting, etc., which represent users' interest in the respective learning objects browsed, tagged, or commented on by the users. These datasets provide too few explicit preference values for example in form of five-star ratings. In general, users are less likely to show their interest in an object by giving explicit ratings. Instead, we can extract their implicit interest in an object by monitoring their activities within a social online platform like the one for ODS. Some of the similarity measures used in the Collaborative Filtering (CF) algorithms such as Pearson correlation, and Cosine are not suitable choices for this kind of data because they require explicit user preference values for measuring similarity between users. We chose Tanimoto-Jaccard coefficient and Loglikelihood ratio since they can deal with implicit users interests in forms of the binary data [1–3].

Besides, we used a graph-based algorithm called T-index [4] that has been mainly designed to improve prediction accuracy of the generated recommendations even when the user data is sparse that is often the case in the educational domain [1]. The original version of the T-index is only based on ratings data of users. Since we want to consider additional social data of users, we extended the T-index to be able to process this kind of users data as well. We tested the extended version of T-index on the MACE, OpenScout and MovieLens datasets.

3 Outcome

Based on the results we have achieved so far, the extended T-index algorithm provides a steady pattern based on F1 score when size of neighbors (n) increases. As mentioned earlier, MACE and OpenScout contain the most similar data to the ODS future data. For these datasets, although Jaccard-Tanimoto provides better F1 score only for a specific size of neighbors e.g. for n=7 in case of MACE, T-index steadily outperforms the used classical CF algorithm. The classical CF algorithms only perform well when enough ratings data are available, as they need a user-item matrix for generating recommendations. The T-index recommender, however, generates recommendations by traversing graphs of users and works well even when the ratings data are sparse. In addition to the performance results, we also showed that T-index helps us to have a balance distribution of degree centrality that provides users with more opportunities for finding central users. The results will be presented in the demonstration session of the conference.

Reference

[1] K. Verbert, H. Drachsler, N. Manouselis, M. Wolpers, R. Vuorikari, and E. Duval, "Dataset-driven research for improving recommender systems for learning," in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, 2011, pp. 44–53.

[2] C. Cechinel, S. Sicilia, Miguel-Ángel Sánchez-Alonso, and E. García-Barriocanal, "Evaluating collaborative filtering recommendations inside large learning object repositories," *Information Processing & Management*, 2012.

[3] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Analysis of Recommendation Algorithms for E-Commerce," in *Proceedings of the 2nd ACM conference on Electronic commerce*, 2000, pp. 158–167.

[4] S. Fazeli, A. Zarghami, N. Dokoohaki, and M. Matskin, "Elevating Prediction Accuracy in Trust-aware Collaborative Filtering Recommenders through T-index Metric and TopTrustee lists," *JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE*, vol. 2, no. 4, pp. 300–309, 2010.

[4] http://movielens.umn.edu