MoocRec.com : Massive Open Online Courses Recommender System

Panagiotis Symeonidis
Department of Informatics
Aristotle University
Thessaloniki, 54124, Greece
psymeon@gmail.com

Dimitrios Malakoudis
Department of Informatics
Aristotle University
Thessaloniki, 54124, Greece
dmalakoudis@gmail.com

ABSTRACT

Massive open online courses (MOOCs) have recently gained a huge users' attention on the Web. They are considered as a highly promising form of teaching from leading universities such as Stanford and Berkeley. MoocRec.com is a web site that recommends courses to users so that, they can acquire those skills, that are expected from their ideal job posting. MoocRec's recommendation engine is based on Matrix Factorization (MF) model combined with Collaborative Filtering (CF) algorithm, which exploits information from external resources (i.e., users' skills, courses' characteristics, etc.) to predict course trends and to perform rating predictions according to them.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information Filtering

1. INTRODUCTION

Massive Open Online Courses (MOOCs) platforms offer thousands of different courses and each course's registration/enrolment can be in the hundreds of thousand students. It would be very useful, if someone could be recommended a course to acquire those skills, that are expected from his ideal job description.

MoocRec.com is a web site that provides to users recommendations of MOOCs. Firstly, users provide some information about their studies and their dream job. Then, MoocRec.com recommends to them related courses, to acquire the required skills for getting their dream job. The heart of the recommendation engine of MoocRec.com is matrix decomposition over a user-course rating matrix R to reduce its dimensions and remove noise from data. To do this, we preserve a small number of k latent features (i.e., dimensions) with the objective to reveal the mainstream users' preferences. For example, in Figure 1, we plot users and courses, assuming that k has been tuned to 2.

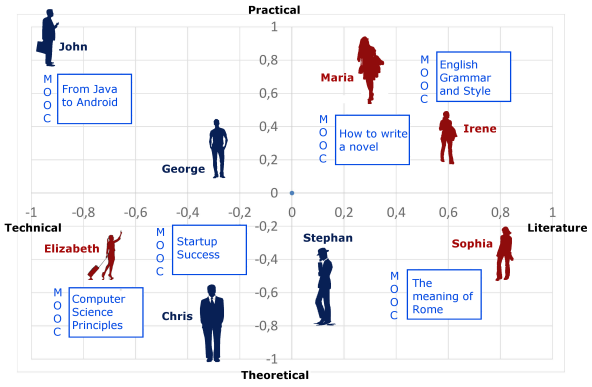

Figure 1: Users and Courses in the 2-D space.

As shown in Figure 1, courses/users that are placed in close distance, are the most suitable/similar to each other. As shown, women prefer literature courses, whereas men choose the technical ones. Specifically, the course "English Grammar and Style" can be recommended to Maria and Irene, whereas "From Java to Android" course is more suitable to John. Please notice that matrix decomposition has also revealed a second separation, which takes place among people's preference, towards practical and theoretical types of courses. In MoocRec.com, we predict users' ratings over courses based on matrix factorization (MF) technique, which exploits information from several external resources/matrices.

The rest of this paper is organized as follows. Section 2 summarizes the related work, Section 3 summarizes the system's architecture. Section 4 describes our recommendation engine. Finally, Section 5 concludes this paper.

2. RELATED WORK

Furnas et al. [1] proposed Singular Value Decomposition (SVD) in Information Retrieval research field. More specifically, SVD captures latent associations between the terms and the documents. SVD is a well-known factorization technique that factors a matrix into three matrices. An instance of SVD, known as UV-decomposition, searches for two matrices (U and V), whose their multiplication gives an approximation of the original matrix R. A significant improvement on the prediction accuracy of classic MF algorithm may be obtained through the incorporation of implicit feedback into the MF model [2, 3, 4].

Extensions of classic MF algorithm have been successfully applied for recommendations in MOOC domain to address the problem of high students' drop-out rates from online courses. To reduce the high students' drop-out rates, they provide recommendations of useful forum threads to students based on their blog history inside a MOOC discussion forum. For instance, Yang et al. [5] have designed a latent feature model to describe student behaviors inside a MOOC forum.

3. SYSTEM'S ARCHITECTURE

Figure 2 introduces the architecture of the MoocRec system. It consists of the web site, the recommendation engine, the database and the web crawler.

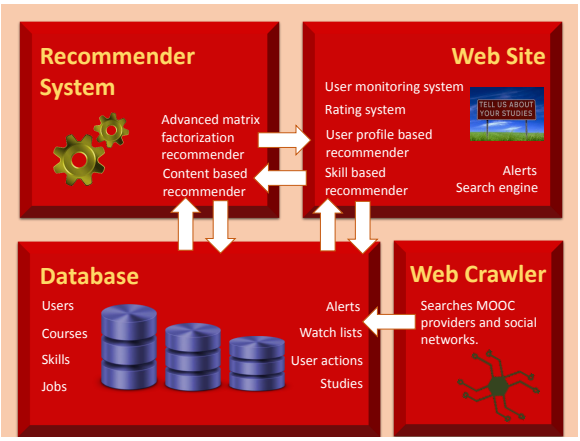

Figure 2: MoocRec System Architecture

MoocRec.com integrates a search engine for MOOCs that are automatically retrieved using web content mining techniques from MOOC providers such as edX and Coursera. It also incorporates a MOOC recommender system, which provides a target user with personalized content according to the skills he wishes to acquire. Therefore, we use content-based filtering in order to connect course's description with the desired skills. We also embed the matrix factorization technique and recommend MOOCs to a target user based on previous choices of other users.

Moreover, using alerts a registered user may be notified of MOOCs when they become available. He may also rate courses either explicitly or implicitly (log history). That is, our monitoring system records the user's actions and helps in constructing his profile.

The web crawler scans edX and Coursera web sites and enters all the MOOCs found in a MySQL database, In addition, an additional web crawler mines from LinkedIn the skills which are connected to the top-rated jobs. Therefore, the database holds information that correlates jobs to skills and skills to courses, which makes course and job recommendation possible.

4. RECOMMENDATION ENGINE

Figure 3 illustrates an example of the course recommendations which are provided by our system. An important characteristic of our recommendations is that the user is informed about the reason he was recommended a course. For example, course "Data Analysis for Life Sciences 6: High-performance Computing for Reproducible Genomics" is recommended because it is related to Life Sciences and Genomics. Our recommendation algorithm, places first the courses which provide the greatest number of skills to the target user. Moreover, courses that start soon or are self-paced have priority towards others.

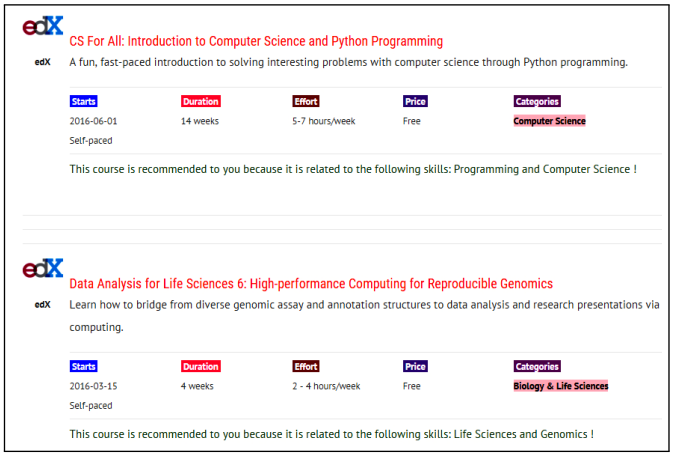

Figure 3: Recommendation appearance and reasoning.

5. CONCLUSIONS

In this paper, we proposed MoocRec.com, which exploits information from external resources (i.e., users' skills, courses' characteristics, etc.) to provide course recommendations. In future, we want to test experimentally our system to check its accuracy effectiveness in terms of accurate recommendations.

References

[1] G. Furnas, S. Deerwester, and S. et al. Dumais. Information retrieval using a singular value decomposition model of latent semantic structure. In *Proc. ACM SIGIR Conf.*, pages 465–480, 1988.

[2] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.

[3] Yehuda Koren and Robert M. Bell. Advances in Collaborative Filtering. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, *Recommender Systems Handbook*, pages 145–186. Springer, 2011.

[4] Arkadiusz Paterek. Improving regularized singular value decomposition for collaborative filtering. In *Proceedings of KDD cup and workshop*, volume 2007, pages 5–8, 2007.

[5] Diyi Yang, Mario Piergallini, Iris Howley, and Carolyn Rose. Forum thread recommendation for massive open online courses. In *Proceedings of 7th International Conference on Educational Data Mining*, 2014.