

SRS: A Subject Recommender System to Enhance E-learning Personalization

Jie Lu*, Xuetao Guo*, Anh Thu Huynh**, Laurie Benkovich*

**Faculty of Information Technology
University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia
Email: {jielu, xguo, laurie}@it.uts.edu.au*

***Faculty of Engineering
University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia*

Abstract. E-learning personalization focuses on how to response individual learner's interests and needs in an e-learning environment. Recommender system technique can effectively implement e-learning personalization. This study develops a subject recommender system (SRS) aimed to locate right subject information to right students according to their individual interests and needs in subject selection. With a technical advance, the system integrates content-based filtering approach, collaborative filtering approach and web usage analysis technique.

Keywords: e-learning, personalization, recommender systems, web usage mining

1. Introduction

E-learning has received much attention in the academic and educational communities. According to the analysis of International Data Corporation (<http://www.idp.com>), the worldwide corporate e-learning market will exceed US\$ 24 billion by 2004. The reason for the growth of is that e-learning provides a convenient and efficient learning environment and practical utilities at anytime and anywhere. Many universities, corporations and educational organizations have developed or are developing web-based learning platforms to provide various course materials and information for online learners. Web personalization is a new direction in online service. It focuses on how to response individual user's interests and needs. Currently, e-learning is much more than putting course information online. Personalized service is needed in e-learning to help online learners learn more efficiently. E-learning personalization proposed in the study attempts applying web personalization techniques into e-learning.

Recommender system techniques have been developed to implement personalised online service in the applications of e-commerce [1]. A recommender system helps users to discover information in relevance to their particular area of interest. Such a system suggests information based on user's preference and past behaviour. Many kinds of approaches and techniques for recommender systems have been developed that utilize either demographic, content, or historical information [2]. Two main types adopted in recommender systems are content-based approach and collaborative filtering (CF) approach [3]. Content-based approach tries to recommend items similar to those a given user has liked in the past. CF identifies users whose tastes are similar to those of the given user and recommends items they have liked. It has been used in a number of different applications such as recommending web pages, movies, articles and products [4]. Recent studies have suggested web usage mining as an enabler to overcome the problems associated with CF since it will reduce the need for obtaining subjective user ratings or registration-based personal preferences [5].

Web usage mining is the process of applying data mining techniques to the discovery of behavior patterns based on web click-stream data. In the advance of e-commerce, the importance of web usage mining grows larger than before [2, 6]. The main tasks of web usage mining generally focus on twofold: data preparation and data pattern discovery. The data preparation tasks build a server session file where each session is a sequence of requests of different types made by single user during a single visit to a site. The pattern discovery tasks involve the discovery of association rules, sequential patterns, usage clusters, page clusters,

user classifications or any other pattern discovery method [2]. As web usage data provide essential information to understanding user preference, through analyzing such click-stream data, it is possible to make more accurate knowing of user's interest. Therefore, usage mining based recommendation approach will take more advantages [7].

This study attempts applying recommendation techniques into e-learning environment to implement e-learning personalization. This study develops a subject recommender system (SRS) prototype in an education environment in the context of the University of Technology Sydney. The system uses both content-based filtering and collaborative filtering approaches by combing with web usage analysis technique. The main function of the system is to locate right subject information to right students according to their individual needs and interests.

2. Recommendation Approach used in SRS

The SRS aims to advise the subjects based on the students' needs and preference. The SRS adopts existing hybrid recommendation technique and adds web usage analysis to overcome the drawback of current technique in the context of subject selection. The basic idea used in the design of the SRS is that the search function first matches subjects that are associated with student query, the content-based technique is used to recommend other relevant subjects, and then, what other students with similar tastes are also interested in are recommended based on CF technique.

To fulfill a content-based scenario, the system expands the scope of the search based on the keyword entered. The subjects that are returned by the keyword search will have other keywords associated with them. This set of new keywords generates a broader search for other associated subjects. The system recommends the second set of subjects as 'subjects the user may also like'.

To facilitate the CF scenario, the past students logs are analyzed for other students who searched on the same particular keywords. The system then generates another search based on other keywords that the students have searched on previously. The system returns the search results as subjects 'other students have also been interested in'. To overcome the sparsity problem associated with most CF techniques, we propose novel two-level CF in the SRS. If the first-level CF fails to produce a result, the second-level will be invoked. This may increase the likelihood that there will be a recommendation based on the CF technique given to the student. The subject set output of the content-based recommender is available at the start of the CF recommender, as well as the keyword that was inputted by the student. The first-level CF attempts to recommend subjects based on a student's interests described by the keyword entered. The CF recommender parses previous searches for the student that has searched on the same keyword as the current student the most, and based on the idea that this past student has highest similarity to the current student, the recommender system suggests other subjects this student has enrolled in. Any new subjects found will be recommended as 'another student who liked your keyword and enrolled in' these subjects. If the past student has not enrolled in any subjects, or no new subject surfaces, the first-level CF returns no result and will invoke the second-level CF. The second-level CF takes the subject set returned from the content-based recommender. The recommender will parse the historical data of previous students, and finds all students that have searched on the same particular keyword as the current student. The recommender recommends the new subjects that the larger set of neighbourhood students have enrolled in. The set of subjects returned though this algorithm is recommended to the student as 'other students who searched on your keyword also liked' these subjects.

To generate a precise recommendation, the most important step is that of computing the similarity between students as it is used to form a proximity-based neighborhood between a target student and a number of like-minded students. The proximity between two students is usually measured using either the correlation or the cosine measure [1]. The main goal of

neighborhood formation is to find, for each student u , an ordered list of l students $N = \{N_1, N_2, \dots, N_l\}$ such that $u \notin N$ and $sim(u, N_1)$ is maximum, $sim(u, N_2)$ is the next maximum and so on. In this study, we adopt cosine measure method [13]. When two students a and b are thought of as two vectors in the m dimensional subject space. The proximity between them is measured by computing the cosine of the angle between the two vectors, which is given by

$$sim(a, b) = \cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|_2 \cdot \|\vec{b}\|_2} = \frac{\sum_{k=1}^m r_{ak} \times r_{bk}}{\sqrt{\sum_{k=1}^m r_{ak}^2} \sqrt{\sum_{k=1}^m r_{bk}^2}}$$

where a and b are row vectors of scale r that student most-liked subjects based on the keywords entry frequency and their subject enrolment priority for two students a and b , respectively. The "." denotes the dot-product of two vectors. Using the similarity measure, for n customers, an $n \times n$ similarity matrix is computed. The $sim(a, b)$ value range from 0 to 1, where more similar vectors result in bigger value.

3. SRS Design and Implementation

The SRS relies on the keywords attached to each subject to make its recommendations to the student. Each subject is tagged by keywords. Once logged in, the student will be able to search for subjects based on a keyword that is entered into the system. If the keyword is within the system, the search will be saved, and subjects that are returned by the keyword search will be displayed. The system will then go on to display up to two sets of recommended subjects that the student may like. The first set of recommended subjects will originate from the content based filtering technique, the second from the CF approach.

The SRS is comprised of five subsystems: Student, Enroll, Subject, Search, and Recommender. The student subsystem is responsible for storing information and accessing information about an individual student. The enroll subsystem is responsible for storing information and accessing information about student enrolment. The subject subsystem is responsible for storing information and accessing information about keyword-subject associations. The search subsystem is responsible for storing information on searches. The recommender subsystem is to draws information about searches, subjects, and students and provides a data mining facility to retrieve a subject for a student.

In the SRS, subjects are tagged against keywords. A keyword is entered into system as input to a search engine for subjects. Content-based filtering uses the keywords tagged against the subjects returned to recommend other subjects. CF uses the stored searches of other students based on the same keyword to recommend other subjects. Referring to the starting point of the content-based recommendation is the returned subjects of the search. The system will find all keywords that are associated with the subjects returned (not including the keyword inputted by the student). If there are no new keywords found the content-based recommender will not give any subject recommendations. If there are new keywords found, the recommender will then search for all subjects associated with the new keyword set having a priority higher than that of 5, number 1 being the highest priority association. If there are any new subjects found, the recommender will recommend these subjects as 'the student may like'. This new larger subject set will be sent to the starting point of the CF recommender.

The SRS has been implemented at <http://charlie.it.uts.edu.au/~athuynh/form1.html>, and an example of recommendation result is shown in Fig. 1. The system is scripted using PHP as PHP is HTML-embedded scripting language that can be used to develop interactive web applications. PHP also works well with databases MySQL which is used in the SRS.

The SRS offer advantages as follows: maintainers of the SRS are able to control the content-based recommendation to some extent by prioritizing the keyword to subject

associations. Records are updated 'on the spot' so that notes of interest and enrolment updates are immediately available as inputs to the CF algorithms of the SRS. Once a student has some subjects displayed, either through the search engine, or by means of a recommendation, the student can click on the subject code to view a subject of interest. This action will register that that particular student was interested in that particular subject. The SRS has two levels of interest available: the first level is the interest-factor, while the second level of interest is at the enrolment of the subjects. This makes the recommendation of subjects has a high acceptance by students.

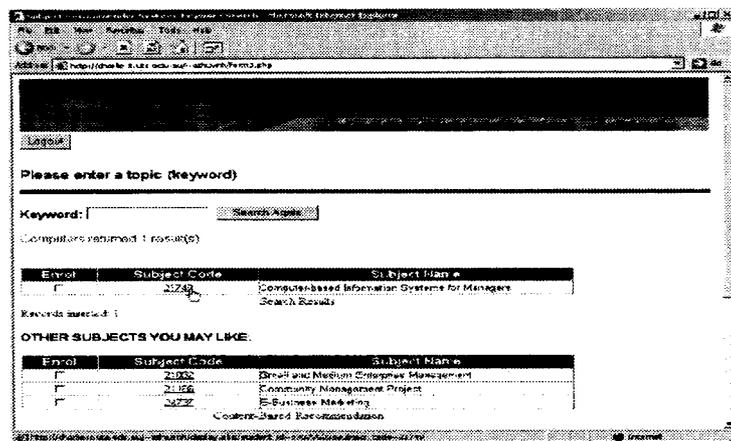


Fig. 1. Example screen of recommendation result

4. Conclusion and Further Study

In this study, we have developed and implemented a subject recommender system with a novel approach integrated content-based filtering, CF and web usage analysis technique. The two-level CF technique proposed in the system attempts to implement e-learning personalization and overcome sparsity problem. However, as a preliminary study, the system has several limitations that we need investigate further: (1) the system can only be used by students who already registered in the university. This was done as it was assumed only students enrolled at university could access the SRS system; (2) the keyword entered by students has to match keywords exactly as they exist in the SRS. There is no handling of possible misspellings or suggestions of similar keywords; and (3) enrolment of subjects is only for one set of subjects. There is no distinguishing feature to distinguish between semesters or years. The further work of the study will conduct an experiment to validate and evaluate the system performance.

References

- [1] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for E-commerce," presented at 2nd ACM E-Commerce 2000 Conference, Minneapolis, Minnesota, 2000.
- [2] M. Balabanovic and Y. Shoham, "Fab: content-based collaborative recommendation," *Communications of the ACM*, vol. 40, pp. 66 - 72, 1997.
- [3] G. Linden, B. Smith, and J. York, "Amazon.com recommendation: item-to-item collaborative filtering," *IEEE Internet Computing*, 2003.
- [4] P. Resnick and H. R. Varian, "Special issue on recommender systems," *Communications of the ACM*, vol. 40, pp. 56 -58, 1997.
- [5] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic personalization based on web usage mining," *Communications of the ACM*, vol. 43, pp. 142-151, 2000.
- [6] R. Cooley, B. Mobasher, and J. Srivastava, "Data preparation for mining world wide web browsing patterns," *Journal of Knowledge and Information Systems*, vol. 1, 1999.
- [7] J. Srivastava, R. Cooley, M. Deshpande, and P. Tan, "Web usage mining: discovery and applications of usage from web data," *SIGKDD Explorations*, vol. 1, pp. 1-12, 2000.

ISBN 4-901329-02-2

Information

Proceedings of the Third International Conference on



November 29 - December 2, 2004,
Hosei University, Tokyo, Japan

Edited by
Lei Li
Kang K. Yen

Published by International Information Institute
www.information-iii.org

02
03
04
05
06
07
08
09
10
12
13
14
16
18
19
20
21
22
23
24
25
26
27
28
29

