Using data mining in a recommender system to search for learning objects in repositories

A. ZAPATA-GONZALEZ, Autonomous University of Yucatán, México
V.H. MENENDEZ, Autonomous University of Yucatán, México
M.E. PRIETO-MÉNDEZ, University of Castilla la Mancha, Spain AND
C. ROMERO, University of Cordoba, Spain

This paper describes a hybrid recommender system for personalizing the search for Learning Objects (LOs) by using data mining. In fact, a nearest-neighbors based approach has been used to discover the most similar LOs and users, and rule mining used to discover LOs that have been downloaded together by similar users.

Key Words and Phrases: Learning Objects, Hybrid Recommender Systems, Association rule mining.

1. INTRODUCTION AND DESCRIPTION OF THE RECOMMENDER SYSTEM

This paper proposes a hybrid recommendation method to assist users (normally instructors) in personalized searches for Learning Objects (LOs) in repositories. The method proposed (see Figure 1) provides a ranking of LOs, starting from a query by applying the following three phases: preselecting LOs, applying filters and ranking LOs.

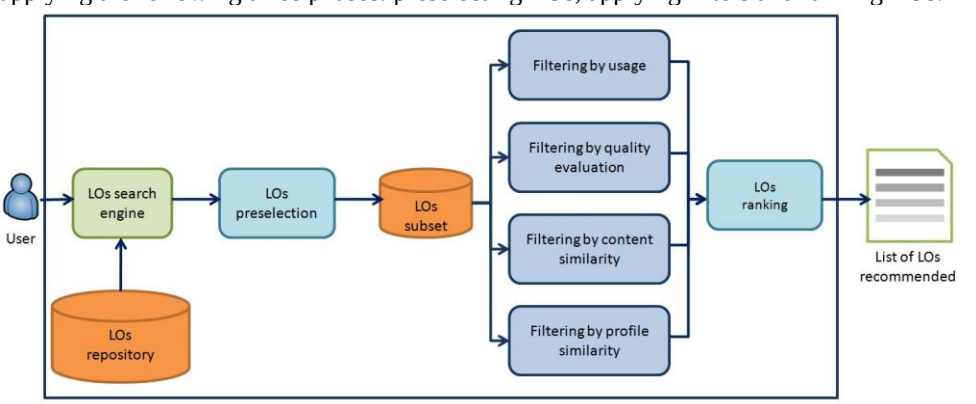

Fig. 1. Proposed architecture for the LO hybrid recommender method.

Firstly, a user does a query using the search engine based on keywords and/or values of some relevant metadata associated with LOs in order to preselect only a subset of LOs that contain the desired contents. Then four filtering or recommendation techniques are applied in parallel: filter by usage (with a ranking of the most commonly downloaded LOs), filter by evaluation (to rank according to the best LOs evaluated), filter by content similarity (ranked according to the most similar LOs) and filter by profile similarity (according to the most similar users/authors). The last two filters use the Nearest Neighbors approach (Ricci et al., 2011) that is like a lazy leaner classification algorithm. For a given LO or user, it compares the LO's or user's attributes to the rest of the LO

Authors' addresses: A. Zapata-Gonzalez, V.H. Menendez, Autonomous University of Yucatán, Mexico, E-mail: zgonzal@uady.mx, mdoming@uady.mx; M.E. Prieto-Méndez, University of Castilla la Mancha, Department of Computer Science, Spain, E-mail: Manuel.Prieto@uclm.es; C. Romero, Department of Computer Science, University of Cordoba, Spain, E-mail: cromero@uco.es;

subsets or their authors by measuring their similarity. This metric lets us determine the degree of similarity between two LOs (by comparing metadata values) or between two users (by comparing values in profile attributes). Then, the outputs (partially ordered LO sets) of the four filters are combined in the final LO ranking. This ranking is obtained by a weighted composition in order to be able to attach more or less importance to one filter or another. This hybrid recommender method has been implemented into the AGORA repository (Prieto et al., 2008) and is called DELPHOS. Figure 2 shows an example of a ranking of LOs obtained by DELPHOS.

Fig. 2. An example of search results found by DELPHOS.

For each LO recommended, DELPHOS shows the user the following information (see Figure 2): type, full name and description of the LO, rating (by using a five start scale as graphical representations of the numerical value), some statistics (total number of downloads, visualizations and evaluations), a brief explanation about why this particular object has been recommended (by using the values obtained in each filter), a list of related users (who have rated/evaluated the LO highly) and a list of related LOs (that have been downloaded together). Association rule mining has been applied to reveal these relationships (or item-item correlations) among current recommended LO and other LOs which have often been downloaded together by users like the one who is currently carrying out the search. In fact, Apriori Predictive (Scheffer, 2005) is used because it is a parameter-free association algorithm that does not require the user to specify either the minimum support threshold or confidence values and is therefore easier for a non-expert in data mining to manipulate. Finally, the user can visualize or download one or several of the lists of recommended LOs (and provide implicit information) and then he/she can rate them using a questionnaire (and provide explicit information). All the information provided can be used to evaluate the accuracy of the recommendations.

ACKNOWLEDGEMENTS

This research has been partially supported by the Regional Government of Andalucia P08-TIC-3720 project; TIN2010-20395 FIDELIO project, MEC-FEDER, Spain, PEIC09-0196-3018 SCAIWEB-2 excellence project, JCCM, Spain and POII10-0133-3516 PLINIO project, JCCM, Spain; The National Council of Science and Technology (CONACYT, México) and the Council of Science and Technology of Yucatán State (CONCyTEY, México).

REFERENCES

PRIETO, M. E., MENENDEZ, V., SEGURA, A. VIDAL, C. 2008. A Recommender System Architecture for Instructional Engineering. *Emerging technologies and Information systems for knowledge society*, 314-321.

RICCI, F., ROKACH, L., SHAPIRA, B., KANTOR, P.B. 2011. Recommender Systems Handbook. Springer.

SCHEFFER, T. 2005. Finding association rules that trade support optimally against confidence. *Intelligent Data Analysis*, 9(4), 381–395.