Moreno, M. N., García-Peñalvo, F. J., & Polo, M. J. (2004). An architecture for personalized systems based on web mining agents. In N. Koch, P. Fraternali, & M. Wirsing (Eds.), Web Engineering. 4th International Conference, ICWE 2004, Proceedings (Munich, Germany, July 26-30, 2004) (pp. 563-567). Berlin: Springer Verlag.
doi:10.1007/978-3-540-27834-4_68

An Architecture for Personalized Systems Based on Web Mining Agents

María N. Moreno*, Francisco J. García and M. José Polo

Dept. Informática y Automática. University of Salamanca. Salamanca. Spain
*E-mail: mmg@usal.es

Abstract. The development of the present web systems is becoming a complex activity due to the need to integrate the last technologies in order to make more efficient and competitive applications. Endowing systems with personalized recommendation procedures contributes to achieve these objectives. In this paper, a web mining method for personalization is proposed. It uses the information already available from other users to discover patterns that are used later for making recommendations. The work deals with the problem of introducing new information items and new users who do not have a profile. We propose an architectural design of intelligent data mining agents for the system implementation.

1 Introduction

The task of finding products or services for Internet users is becoming more tedious every time. The numerous web sites existing nowadays make available more information than a user can manage. On the other hand, the quick growth of the electronic business activities has contributed to increase the market competition. A way for improving the competitiveness of traditional trade companies is to take advantage of business intelligence strategies supported by techniques like data mining. In the e-commerce environment these procedures can also be applied but they have been extended to deal with problems of the web systems. Personalized recommender systems provide users with intelligent mechanisms to search products to purchase. This is a way to avoid the problem of *information overload* due to the great quantity of information accessible through the Web [1].

The quality of the recommendations for the users has an important effect on the clients' retention. Users refuse poor recommender systems which can cause two types of error: *false negatives*, which are products that are not recommended, though the customer would like them, and *false positives*, which are products that are recommended, though the customer does not like them [2]. The most serious errors are false positives, because these errors will cause negative reactions in the customers and thus they won't probably visit the site again. The use of procedures to find customers characteristics that increase the probability of buying recommended products can help to avoid these problems. Data mining techniques and intelligent agents play an important role in the development of efficient personalized recommender systems. In this work, we present a recommendation methodology

based on web mining that uses diverse information as user's attributes, rating and usage data. The core of the methodology is an algorithm that generates and refines association rules used for making personalized recommendations. Our proposal should provide recommender systems with more relevant patterns that minimize the recommendation errors. The architecture suggested for these systems is constituted by intelligent agents; one of them is in charge of doing data mining tasks.

2. Related work

The two main recommendation methods are: collaborative filtering and a content-based approach [4]. The first technique is one of the most successful methods and it was based initially on nearest neighbor algorithms. These algorithms predict product preferences for a user, based on the opinions of other users. The opinions can be obtained explicitly from the users as a rating score or by using some implicit measures from purchase records as timing logs [12]. In the content based approach text documents are recommended by comparing between their contents and user profiles. The main shortcoming of this approach in the e-commerce application domain, is the lack of mechanisms to manage web objects such as motion pictures, images, music, etc. Collaborative filtering also has limitations in the e-commerce environment. Rating schemes can only be applied to homogeneous domain information. Besides, sparsity and scalability are serious weaknesses which would lead to poor recommendations [2]. Sparsity is caused by the fact that the number of ratings needed for prediction is greater than the number of the ratings obtained. The reason for this is that collaborative filtering usually requires user explicit expression of personal preferences for products. Performance problems in searching for neighbors is also a limitation. The computation time grows linearly with both the number of customers and the number of products in the site. Another obstacle is the first-rater problem that takes place when new products are introduced [3].

There are two approaches for collaborative filtering, *memory-based* (*user-based)* and *model-based* (*item-based)* algorithms. ***Memory-based*** algorithms, also known as *nearest-neighbor* methods, were the earliest used [11]. They treat all user items by means of statistical techniques in order to find users with similar preferences (*neighbors*). The advantage of these algorithms is the quick incorporation of the most recent information, although the search for neighbors in large databases is slow [13].

Data mining technologies have also been applied to recommender systems. ***Model-based*** collaborative filtering algorithms use these methods in the development of a model of user ratings. This approach was introduced to reduce the sparsity problem and to get better recommender systems. Some examples of these methods are the Bayesian network analysis [13], the latent class model [1], rule-based approaches [4], decision tree induction combined with association rules [2], horting [14]. Web mining methods build models based mainly on users' behaviour more than in subjective valuations (ratings). This is the main advantage of this approach that allows avoiding the problems associated with traditional collaborative filtering techniques [6].

3 Recommendation methodology

In this section, a methodology for personalized recommendations is presented. It uses information about consumer preferences and user attributes.

The first step of the methodology is the selection of the best rated products. A list of products ordered from most to less popular is generated. The aim of this step is to reduce the number of association rules generated and to obtain rules applicable to a wide range of customers. The selected records are the inputs for the second step in which the association rules are generated. The rules relate product attributes with user attributes and preferences, in this way it is possible to identify products that will be of interest to a given customer with a specific profile. The initial rules are refined in order to obtain strong patterns that avoid the false positive recommendations. The association rules are also a solution to the problem of the introduction of new users. When a new customer uses the system his profile is obtained and recommendations for him are generated according his profile. In the third step, the recommendations are made. The recommendations are based on the patterns obtained, which relate user attributes with product attributes. For new and old users the system recommends products with characteristics adapted to their profile. The rules enable to recommend new products whose characteristics agree with the preferences of the users. New products with new characteristics are always recommended. This is the way to deal with the first-rater problem. The process is iterative and uses the new information about products and users to feedback the system. New association models are built when the system has a significant quantity of new information. The architectural design of the system is shown in figure 1. It contains three main agents:

- Data mining agent builds the recommendation models. It is in charge of generating and refining association rules. It uses the information provided by the data management agent periodically.
- Recommendation agent receives the requests from the users, takes their preference profile and uses the data mining models to make personalized recommendations.
- Data management agent collects and manages the storage of the information about new user preferences and new products. It is connected with the data mining agent to provide the data that are used periodically in the generation of new models.

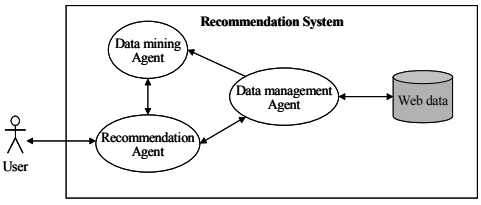

Figure 1. Simplified architecture of the recommender system

4 Association analysis

Methods for rule discovery are widely used in many domains. They have been adopted to target marketing or personalized e-commerce recommendation service.

We use the concept of unexpectedness [9] for refining association rules. The refinement of association rules provides recommender systems with more confident

rules and serves to solve conflicts between rules. In an earlier work [7] we have used several visualization techniques and data mining methods to build and validate models for software size prediction. We found that the best attributes for classification give good results in the refinement of associations rules in the area of projects management. The presented recommender procedure follows the approach of using these attributes in a rules' refinement algorithm wich works with web usage data. In classification problems, the *label* attribute is the target of the prediction process. The attributes used are the best in discriminating the different values of the label attribute. They are obtained by computing the entropy of the attribute [8].

The refinement procedure starts with a set of rules which relate items. Then we search for unexpected patterns that could help us to increase the confidence or to solve ambiguities or inconsistencies between these rules. The refinement process fits into a generic iterative strategy [10]. Each iteration consists of three steps: 1) generation of unexpected patterns for a belief, 2) selection of a subset of unexpected patterns that will be used to refine the belief, and 3) refining the belief using selected patterns and the best attributes for classification. The process ends when no more unexpected patterns can be generated. We have instantiated this generic strategy and created a specific refinement process [8]. The principal feature of our approach is the gradual generation of the unexpected patterns. We take advantage of knowledge of good attributes for classification and use them progressively, beginning with the best. This simplifies the selection of patterns and the refinement process.

5 Experimental data treatment

The experimental study was carried out using data from MovieLens recommender system developed by. the GroupLens Research Project at the University of Minnesota The database contains user rating about movies that belong to 18 different genres.

Association rules were produced by taking the records with a rate value greater than 2.Initial rules were generated and visualized by using *Mineset* [5]. Figures 2 and 3 are graphical representations of first and refined rules respectively (LHS → RHS).

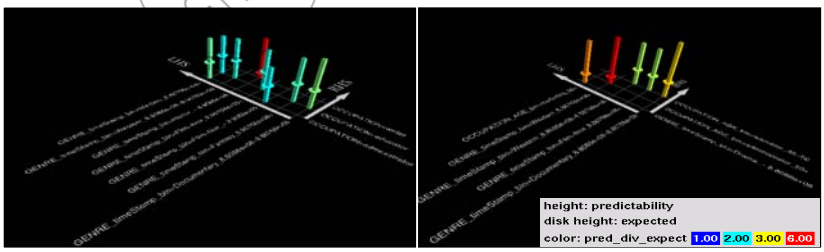

Figure 2. Initial rules **figure 3**. Refined rules

Initial rules represented in the figure 2 are rules that relate "Genre-Time Stamp" with the user's occupation. Time stamp is the time that the user spent in a product. Refined rules (figure 3) use the attribute "age" combined with "occupation" RHS. Good rules are those with high values of pred_div_expect (predictability/expected predictability). The graphs show that these values increase in the refined rules.

6 Conclusions

The personalized recommendation engines contribute to improve the service to the clients and the competitiveness. In this paper, a methodology for recommendation based on an algorithm for refining association rules proposed. The methodology deals with the case of making recommendations for new users. It provides systems with more relevant patterns, which lead to more effective recommendations. Rules obtained from the refinement procedure have higher values of confidence. This means that the recommendations are more accurate and the number of false positive recommendations is reduced.

References

1. Cheung, K.W., Kwok, J.T., Law, M.H. and Tsui, K.C.: Mining customer product ratings for personalized marketing. Decision Support Systems, 35 (2003) 231-243.
2. Cho, H.C., Kim, J.K., Kim, S.H.: A personalized recommender system based on web usage mining and decision tree induction. Expert Systems with App. 23 (2002), 329-342.
3. Konstant, J. Miller, B., Maltz, D., Herlocker, J. Gordon, L. and Riedl, J.: GroupLens: Applying collaborative filtering to usenet news. Comm. of the ACM, 40 (1997), 77-87,.
4. Lee, CH., Kim, Y.H., Rhee, P.K.: Web personalization expert with combining collaborative filtering and association rule mining technique. Expert Systems with Applications 21 (2001), 131-137.
5. Mineset user's guide, v. 007-3214-004, 5/98. Silicon Graphics (1998).
6. Mobasher, B., Cooley, R. and Srivastava, J.: Automatic personalization based on web usage mining, Communications of the ACM, 43 (8) (2000), 142-151.
7. Moreno, M.N., Miguel, L.A., García, F.J., Polo, M.J.: Data mining approaches for early software size estimation. Proc. 3rd ACIS International Conference On Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD'02), 361-368, Madrid, Spain (2002).
8. Moreno, M.N., Miguel, L.A., García, F.J., Polo, M.J.: Building knowledge discovery-driven models for decision support in project management. Dec.Support Syst. (in press).
9. Padmanabhan, B., Tuzhilin, A.: Knowledge refinement based on the discovery of unexpected patterns in data mining. Decision Support Systems 27 (1999) 303– 318.
10. Padmanabhan, B., Tuzhilin, A.: Unexpectedness as a measure of interestingness in knowledge discovery. Decision Support Systems 33 (2002) 309– 321.
11. Resnick, P., Iacovou, N., Suchack, M., Bergstrom, P. and Riedl, J.: Grouplens: An open architecture for collaborative filtering of netnews. Proc. of ACM CSW'94 Conference on Computer.Supported Cooperative Work, 175-186, (1994).
12. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithm. Proc. of the tenth Int. WWW Conference (2001), 285-295.
13. Schafer, J.B., Konstant, J.A. and Riedl, J.: E-commerce recommendation applications. Data Mining and Knowledge Discovery, 5 (2001), 115-153.
14. Wolf, J., Aggarwal, C. Wu, K.L. and Yu, P.: Horting hatches an egg. A new graph-theoretic approach to collaborative filtering. Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, C.A., (1999).