Collaborative Filtering Support for Adaptive Hypermedia

Martin Balík, Ivan Jelínek

Department of Computer Science and Engineering, Faculty of Electrical Engineering
Czech Technical University
Karlovo náměstí 13, 121 35 Prague, Czech Republic
balikm1@fel.cvut.cz

Abstract. Web applications of today are dealing with huge amounts of data. Developers need tools to manage those data efficiently and they need be able to present the most important information to users. Modern applications offer personalization of content and some of the applications are also capable of delivering adapted information based on user needs. One of the possibilities for adapting information presented to the user is collaborative filtering. Information is filtered based on the preferences of similar users. In our work we are developing a general adaptive web model. One of our experiments focused on collaborative filtering algorithms and their application in adaptive systems. We present our approach, experiments and results of our work.

Keywords: adaptive hypermedia, personalization, user modeling, general model, collaborative filtering, algorithms, Web 2.0.

1 Introduction

Collaborative filtering[1] is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints and data sources. We live in the age of information explosion and we need tools to process the large amounts of information and offer users the comfort of dealing only with relevant information. The collaborative filtering techniques proved themselves to be very useful for this. Many existing applications use filtering techniques to recommend items such as music [1], books [2], movies etc. There are many fields where applications benefit from the ability to make qualified recommendations. This could be used even in e-learning to recommend topics appropriate for a user's knowledge [3]. Therefore, we think that such type of information adaptation should be part of the adaptive web framework that we are developing.

In our previous work we proposed a General Ontological Model for Adaptive Environments (GOMAWE) [4]. The collaborative library will be used as part of the reasoning layer (Fig. 1). Using artificial intelligence algorithms we can derive some

[1] http://en.wikipedia.org/wiki/Collaborative_filtering

user characteristics that were not stored in the user model. Similarly, we could add new rules to the multidimensional matrix in the storage layer based on the rules of similar users.

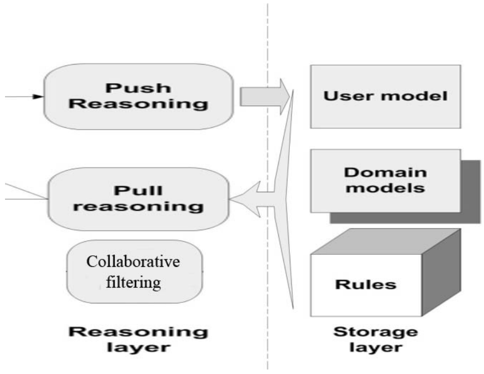


Fig. 1. Collaborative filtering support in GOMAWE

2 Existing applications

Collaborative filtering is based on the assumption that similar users have similar preferences [5]. Recommender systems are typically used in web applications that have many users and want to provide each user with information corresponding to his/her preferences. Amazon.com[2] internet shop portal makes recommendations based on the items users have already bought. Last.fm[3] music portal recommends its users songs based on their recent playlist. Users can also use tags to describe the songs and also to classify them into genre groups. International movie database (IMDb)[4] recommends to the users movies similar to their favorite ones. These are just the most popular portals and the recommendation feature is part of many others.

In adaptive web applications we can also use clustering based on the similarity of users that could help to solve the "cold start" problem. This means that if we don't have any information about the current user, we could use information about a similar user which we assume to be similar too.

There are many approaches and algorithms to compute the similarity of user preferences [6]. We have selected the most simple and most typically used algorithms and we will discuss them in the following section. We have implemented these algorithms as a java program and performed experiments with real data.

---

[2] http://www.amazon.com/
[3] http://www.last.fm/
[4] http://www.imdb.com/

3 Selected algorithms

The commonly used algorithm for collaborative filtering tasks is the **k-Nearest Neighbor** (k-NN) algorithm [7]. The k-NN algorithm is a method for classifying objects based on the closest training examples in the feature space. It belongs to a class of so called lazy learning algorithms. The following formula can be used to calculate the distance d of two users u_a and u_b:

$$d(u_a,u_b)=\sqrt{\sum_{i=1}^{n}(P_a(o_i)-P_b(o_i))^2} \tag{1}$$

where n is the number of compared objects o_i and $P(o_i)$ is the rating of the object. After determining the distance of users, we can select K users with the lowest distance and calculate the unknown rating of item i for the user u_0 as the arithmetic mean of rating of the K nearest users:

$$P_0(o)=\frac{\sum_{i=1}^{K}P_i(o)}{K} \tag{2}$$

We can also use k-NN algorithm to calculate the similarity of two users as a value ranging from 0 to 1 as:

$$sim(u_a,u_b)=1-\frac{d(u_a,u_b)}{\sqrt{n}} \tag{3}$$

Another algorithm for determining the similarity of users uses the **Pearson correlation coefficient** [8]. It ranges from -1 (a perfect negative relationship) to +1 (a perfect positive relationship), with 0 stating that there is no relationship whatsoever. The value of the coefficient can be computed as a quotient of covariance of variables and their standard deviations:

$$r=\frac{\operatorname{cov}(X,Y)}{s_X s_Y}=\frac{E((X_i-E(X))(Y_i-E(Y)))}{s_X s_Y}=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{s_X}\right)\left(\frac{Y_i-\overline{Y}}{s_Y}\right) \tag{4}$$

where $\overline{X}$ and $\overline{Y}$ are sample means and s_x and s_y are sample standard deviations.

Similarity can be also calculated using **Spearman correlation coefficient** [9]. In principle, ρ is simply a special case of the Pearson product-moment coefficient in which two sets of data X_i and Y_i are converted to rankings x_i and y_i before calculating the coefficient. If there are no tied ranks:

$$\neg\exists i,j : i\neq j(X_i=X_j\vee Y_i=Y_j) \tag{5}$$

then ρ is given by:

$$\rho=1-\frac{6\sum_{i=1}^{n}d_i^2}{n(n^2-1)} \tag{6}$$

where d_i is the difference between the ranks of corresponding values X_i and Y_i and n is the number of values in each data set.

The last algorithm that we used for our experiments uses the **Kendall coefficient** [10]. Kendall τ coefficient is defined as:

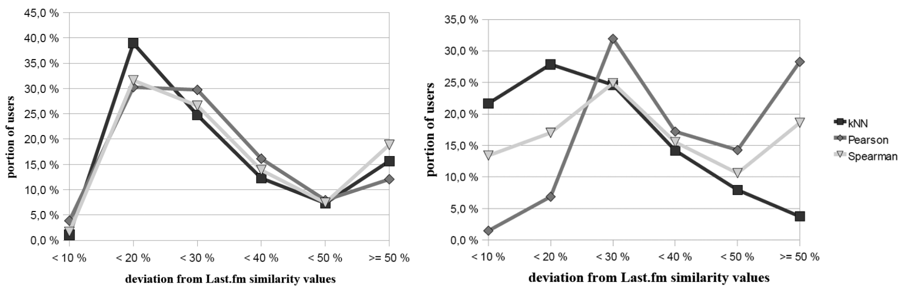
$$\tau=\frac{n_c-n_d}{\frac{1}{2}n(n-1)} \tag{7}$$

where n_c is the number of concordant pairs and n_d is the number of discordant pairs.

4 Experiments

We performed experiments in the field of music recommendation [11]. We used our library with implemented algorithms that we explained in the previous section and we used data from the music portal Last.fm[5]. We have analyzed data of five users and compared the results of our algorithm implementations with the results of the Last.fm service. To be able to compare the data, normalization was needed. All values were multiplied by a coefficient. This coefficient was chosen so that the highest value of similarity computed in our library and the highest value reported by the Last.fm service were equal. All negative values were set to zero. This is because the used input data contained 50 interprets with the best rating from the selected user. Therefore we have no possibility to determine, which items had the worst rating. However, in some other cases even negative values could be used to reason about user preferences.

In Fig. 2. there are the results for two selected users. The values were computed using three of the algorithms and we observed the deviation from Last.fm similarity values. Compared to processing all data, we achieved better results by eliminating users with similarity values near zero.



user “soustruh“ *user ”filipesta”*

Fig. 2. Portion of users depending on the deviation from Last.fm similarity values

We achieved the best results with the k-NN algorithm. The number of processed users has significant influence on the results. Optimalization could be also achieved by changing the *k* value – number of nearest neighbors.

The algorithms were implemented as a java library. For the experiments we used a graphical user interface. However, the library could be used separately, e.g. in the adaptive web portal backend. This will be the next challenge and we will perform experiments corresponding to the scheme of the GOMAWE that we mentioned in the text earlier.

---

[5] http://www.last.fm/

3 Conclusions and future work

We presented our approach to utilize collaborative filtering techniques within adaptive web systems. The algorithms computing the recommendations will be part of the reasoning layer of our General Ontological Model for Adaptive Web Environments (GOMAWE). We performed experiments with three selected algorithms and achieved promising results.

Currently we are developing an adaptive system based on GOMAWE. Our approach allows dealing with adaptation techniques as black box components. Therefore the collaborative filtering library could be used as such an adaptation component. An evaluation of adaptation based on the collaborative filtering library in the system that we are developing will be subject of our future work. Our complete work should lead to developing of a universal framework for adaptive web applications.

Acknowledgments. This research has been supported by MSMT under research program No. 6840770014. This research has been supported by the grant IGS ČVUT No. CTU0915213.
The results of our research are part of the work of a special research group WEBING (http://webing.felk.cvut.cz).

References

1. Pampalk, E., Pohle, T., and Widmer, G.: Dynamic Playlist Generation Based on Skipping Behavior. In Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR'05), London, UK, Sept. 11--15, 2005. 634--637.
2. Linden, G., Smith, B., York, J.: Amazon.com Recommendations. Item-to-Item Collaborative Filtering. IEEE Internet Computing (2003)
3. Miettinen M., Kurhila J., and Tirri H.: On the Prospects of Intelligent Collaborative E-learning Systems. In Proc. 12th International Conference on Artificial Intelligence in Education (AI-ED 2005). IOS Press, pp. 483--490.
4. Balík M., Jelínek I.: Towards Semantic Web-based Adaptive Hypermedia Model. ESWC Ph.D. Symposium, Tenerife, Spain, pp. 1--5 (2008), ISSN: 1613-0073.
5. Grcar M., Mladenic D., and Grobelnik M.: Data quality issues in collaborative filtering. In the proceedings of ESWC05, Heraklion, Greece (2005)
6. Grčar M.: User Profiling: Collaborative filtering. In proceedings of SiKDD 2004, Ljubljana, Slovenia (2004)
7. Nearest-Neighbor Methods in Learning and Vision, edited by Shakhnarovish, Darrell, and Indyk, The MIT Press, 2005, ISBN 0-262-19547-X
8. Moore, David (August 2006). "4". Basic Practice of Statistics (4 ed.). WH Freeman Company. pp. 90–114. ISBN 0-7167-7463-1
9. Myers, Jerome L.; Arnold D. Well (2003). Research Design and Statistical Analysis (second edition ed.). Lawrence Erlbaum. pp. 508. ISBN 0805840370
10. Abdi, H. (2007). [1] (2007) Kendall rank correlation. In N.J. Salkind (Ed.): Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage..
11. Stružský M.: Collaborative filtering for adaptive web. Bachelor theses, Faculty of Electrical Engineering, Czech Technical University in Prague. Czech Republic (2009)