Simulated Analysis of MAUT Collaborative Filtering for Learning Object Recommendation

Nikos Manouselis[1], Riina Vuorikari[2], and Frans Van Assche[2]

[1] Informatics Laboratory, Div. of Informatics, Mathematics & Statistics, Dept. of Science, Agricultural University of Athens, 75 Iera Odos Str., 11855, Athens, Greece
nikosm@ieee.org
[2] European Schoolnet (EUN), 61 Rue de Treves, Belgium
{Riina.Vuorikari, Frans.Van.Assche}@eun.org

Abstract. This paper examines the case of developing a learning resources collaborative filtering service for an online community of teachers in Europe. A data set of multi-attribute evaluations of learning resources has been collected from the teachers who used European Schoolnet's CELEBRATE portal. Using this data set as input, a candidate multi-attribute utility collaborative filtering algorithm is appropriately parameterized and tested for potential implementation on the portal. This simulation experiment may serve as a first step towards the understanding and appropriate specialization of a collaborative filtering service for the given user community.

Keywords: Collaborative filtering, simulation, evaluation, learning object.

1 Introduction

Internet users are often times overwhelmed by the flow of online information, hence the need for adequate systems that will help them manage such situations. Collaborative filtering systems recommend users with items that people with similar preferences liked in the past. A vast number of studies exists that deals with the design, development and evaluation of such systems [2]. Nevertheless, related research identifies that a common drawback of them is that, often, they do not consider the needs of the actual online community that they aim to support [4]. It is rather common to witness evaluation studies of proposed systems that engage usage data from totally different application domains, overlooking the particularities of their own domain [13]. This is why the importance of careful testing and parameterization of collaborative filtering systems, under conditions similar to the ones of their actual application and prior to the actual deployment in real settings, has been already highlighted [4,9].

In the field of Technology-Enhanced Learning (TEL), this is an upcoming reality as well. Numerous new digital learning repositories are set up and users face a plethora of learning resources available online. Thus, online guidance and services to help users identify suitable learning resources from a potentially overwhelming variety of choices could be beneficial. As a consequence, a number of recommender systems that aim at supporting users in finding learning resources online have already been introduced, including several collaborative filtering ones (e.g. [12, 11, 5,3]).

In this direction, this paper describes an experimental investigation of a collaborative filtering service for an existing Web portal, based on authentic data that comes from the online community that uses the portal. The case under study is the CELEBRATE project led by European Schoolnet (EUN), and the user community consists of teachers around Europe. These teachers had access and used the project portal to locate, share, and evaluate learning resources. Teacher-provided evaluations come in the form of multi-attribute ratings, which may be shared among community members, and thus potentially enhance discovery and reuse of the learning resources. More specifically, the aim of this study is to experimentally investigate which variation of a particular algorithm that is based on Multi-Attribute Utility Theory (MAUT) [6] can be implemented to support the collaborative filtering service of the EUN's portal. For this reason, a simulation environment [8] for collaborative filtering algorithms has been used to analyze the authentic teachers' evaluations, to test various parameterizations of the candidate algorithms, and to reach a conclusion about which parameterization is the most appropriate for this online community.

2 Case Study

European Schoolnet (EUN) has carried out work on learning resources interoperability since 1999. This endeavor has resulted in the concept of the Learning Resources Exchange (LRE), a federation of learning resources repositories and portals of European K-12 stakeholders, such as National Educational Authorities, Ministries of Education and private and commercial partnerships. The common aim is to allow European teachers and learners easily locate, use and reuse both open content, as well as content from commercial suppliers.

The online community, whose evaluations were used for this paper, took part in the CELEBRATE project (http://celebrate.eun.org). This project addressed all parts of the educational content value chain and involved 23 participants including Ministries of Education (MeE), universities, educational publishers, content developers, vendors and technology suppliers from eleven European countries. The aims of CELEBRATE were many-fold. On the one hand, there was the interest to find out whether a learning resources' brokerage system, such as the one developed in the project, allows MoEs, publishers and users from individual schools to more easily exchange digital content across national borders, and thus act as a catalyst for the European e-learning content industry. On the other hand, the project aimed to study whether teachers actually like these types of learning resources and if they used them to support their own ICT-based teaching.

In the context of the project, an online K-12 teacher community of about 770 people was formed, with registered members from Finland, France, Hungary, Israel, Norway and the UK. These teachers participated in a large pilot study, one part of which took place in a period of eight weeks from April to June in 2004, when participants accessed about 1,400 learning objects (i.e. digital resources that can be reused to support teaching and learning) and approximately 2,400 assets (i.e. other types of digital resources that could be useful for the teacher community), and provided evaluations about the portal and its content. The objective of the pilot study in general was to contribute to the understanding of teacher's perceived usefulness and quality of the resources they accessed. For this reason, different evaluation methods have been used, combining both online questionnaires that were completed by all community members, as well as interviews with small focus groups.

One outcome of the pilot study is the collection of a large number of multi-attribute evaluations of the resources that the teachers have accessed, viewed and, in most cases, used as well. Teachers were asked to evaluate the learning resources upon several attributes using an evaluation scale between 1-5 [Strongly disagree, Disagree, Neither disagree nor agree, Agree, Strongly agree]. This led to the collection of an evaluations' data set which has been judged as particularly useful for testing a collaborative filtering service a prior to its deployment.. The data set included 2,554 evaluations related to 899 learning resources in CELEBRATE's repositories, which have been provided by 228 teachers. The average number of ratings per user was 11.2.

Data sets with users' feedback, such as the widely known MovieLens and EachMovie data sets [4], are very often used to evaluate collaborative filtering algorithms. However, data sets of multi-attribute evaluations are particularly rare, and usually synthetic (simulated) data sets are being used [13]. Thus we have inclined to use this data set in order to experimentally investigate various design options for a previously proposed collaborative filtering algorithm [9] that takes multi-attribute evaluations as input.

3 Examined Algorithm

In general, the problem of collaborative filtering is to predict how well a user will like an item that he has not rated (also called "evaluated" in the rest of this paper), given a set of historical ratings for this and other items from a community of users [2]. In single-attribute collaborative filtering, the problem space can be formulated as a matrix of users versus items (or user-rating matrix), with each cell storing a user's rating on a specific item. Under this formulation, the problem refers to predicting the values for specific empty cells (i.e. predict a user's rating for an item).

The collaborative filtering problem may be mathematically formulated as it follows [2,7]: let *C* be the set of all users (e.g. the members of an online community) and *S* the set of all possible items that can be recommended (e.g. the digital resources that the community members share and evaluate). A utility function $U^c(s)$ is defined as $U^c(s): C \times S \to \Re^+$ and measures the appropriateness of recommending an item *s* to user *c*. It is assumed that this function is not known for the whole *C* x *S* space but only on some subset of it. Therefore, in the context of recommendation, the goal is for each user $c \in C$ to be able [7]:

- to estimate (or approach) the utility function $U^c(s)$ for an item *s* of the space *S* for which $U^c(s)$ is not yet known; or
- to choose a set of items $S' \subseteq S$ that will maximize $U^c(s)$.

The aim of collaborative filtering is then to predict the utility of items for a particular user (called *active user*), based on the items previously evaluated by other users. That is, the utility $U^a(s)$ of item *s* for the active user *a* is estimated based on the utilities $U^c(s)$ assigned to *s* by those users $c \in C$ who are 'similar' to user *a*.

The goal of CELEBRATE's collaborative filtering service will be to provide some member of the online community (which corresponds to the active user $a \in C$) with an estimation of how he would evaluate a particular target item *s* that he has not previously seen, or with a recommended ranking of items that he has not previously seen, which he would appreciate higher than the others. To calculate this prediction, a neighborhood-based collaborative filtering algorithm is adopted. Neighborhood-based algorithms are the most prevalent approaches for collaborative filtering. The main reason for their popularity is that they are simple and intuitive on a conceptual level, while avoiding the complications of a computationally expensive model-building stage [2]. They have their roots in instance-based learning (IBL) techniques that are very popular in machine learning applications.

The nearest neighbor algorithm is one of the most straightforward IBL algorithms, and uses a function that represents the distance between one instance and another (also called a similarity function) in order to determine how close a new instance is to stored instances. Then, it uses the nearest instance or instances to predict the target. Nearest-neighbor algorithms therefore create a neighborhood $D \subseteq C$ of users that have similar preferences with the active user and who have previously evaluated the target item *s*, and calculate the prediction of $U^a(s)$ according to how the users in the neighborhood have evaluated *s*.

The studied algorithm is a multi-attribute extension of related algorithms, which is based on MAUT [6]. It considers each attribute separately, first trying to predict how the active user would evaluate *s* upon each attribute, and then synthesizing these attribute-based predictions into a *total utility* value. More specifically, we assume that the items are evaluated upon *n* attributes $\{g_1, g_2, \ldots, g_n\}$, where each attribute g_i (i=1,…,n) corresponds on each dimension of an item that is evaluated by a user *c* with an evaluation $g_i^c(s)$. The algorithm creates *n* neighborhoods $D_i \subseteq C$, one for each attribute g_i. The multi-attribute evaluation of an item *s* can be then expressed as a vector $[g_1(s), g_2(s), \ldots, g_n(s)]$. For each attribute, the selection of a user as a potential neighbor is based on the notion of his *similarity* with the active user *a*. In this way, different similarities are calculated for each attribute g_i, and they are denoted as $sim^{g_i}(a,c)$, with i=1,…,n. Using the multiple predictions of the evaluations $g_i^a(s)$ that the active user *a* would give to item *s*, the *n* predictions $g_i^a(s)$ are then used to compute the prediction of the total utility of target item *s*, according to: $U^a(s) = \sum_{i=1}^{n} g_i^a(s)$.

A variety of design options can be considered for the studied algorithm. The main three examined are:

a) Calculation of the similarity between different users using an Euclidian distance measure, a Vector/Cosine measure, and a Pearson correlation factor.
b) Selection of the neighborhood according to a pre-defined Maximum Number of Neighbors or a Correlation Weight Factor.
c) Combining neighbors' ratings into a prediction according to a Simple Arithmetic Mean, a Weighted Mean, or a Deviation from the Mean formula.

A detailed explanation of these options can be found in [9]. Some other multi-attribute algorithm could have been similarly examined [1].

4 Experimental setting

The goal of the experimental testing has been to examine the appropriate parameterization of the proposed algorithm, so that it can be implemented for multi-attribute collaborative filtering of learning resources in the CELEBRATE Web portal. As mentioned before, we have examined a number of design options for the studied algorithm. More specifically, for this experiment, we have considered all three options of Similarity Calculation - that is Euclidian, Vector/Cosine, and Pearson. Both methods for Neighborhood Formation/Selection have been considered - that is Correlation Weight Threshold (CWT) and Maximum Number of Neighbors (MNN). Finally, all three options for Combining Ratings for Prediction have been examined. This led to 3*2*3=18 variations of the proposed algorithm. To fine-tune the algorithm and explore its appropriate parameterization, we further varied the parameter value of the Neighborhood Formation/Selection stage. For CWT, values varied between '0' and '1' (leading to 20 variations). For MNN, values varied between '1' and '20' (leading to 20 variations). The overall number of variations considered have been (20*18+20*18)/2= 360 (from which, 180 using CWT and 180 MNN). The studied algorithm variations have been compared to some basic algorithms, which have been used as standard measures and are explained in [9].

The Collaborative Filtering Simulator (CollaFiS) environment [8] allowed us to parameterize, execute and evaluate all considered variations of the studied algorithm. The data set of multi-attribute evaluations described before has been used as input to the simulator. The evaluations have been processed with CollaFiS, and have been split into one training and into one testing component (using a 80%-20% split). The performance of each algorithm variation has been measured as it follows. For each one of the 511 evaluations in the testing component, the user that had provided this evaluation was considered as the active user, and the evaluated resource as the target item. Then, the algorithm tried to predict the total utility that the target item would have for the active user, based on the information in the training component (2,043 evaluations).

For our experimental testing, three particular performance evaluation metrics have been used: the predictive *accuracy* of the algorithms, through the mean-absolute error (MAE) of the predicted utility against the actual utility of an item; the *coverage*, as the percentage of items for which an algorithm could produce a prediction; and prediction *time*, as the mean time required per item for an algorithm to calculate a prediction and present it to the user. The simulator compared the predicted utility with the actual one, and calculated the MAE from all evaluations in the testing set. Furthermore, it calculated coverage as the percentage of resources in the testing component for which the algorithm could calculate a prediction, based on the data in the training component. Additionally, the time required for a prediction to be calculated and presented to the user has also been recorded. The simulation took place in a PC with a Pentium 4 (2.5GhZ, 256 MB RAM) running Microsoft Windows XP, Apache server 1.3.33, PHP 5.0.3, and MySQL Server 4.1.

5 Results

Figures 1 and 2 present the Accuracy of the various algorithm variations. In both diagrams, the performance of the Euclidian variations is denoted using 'Δ', of the Pearson variations using '*', and of Cosine variations using '□'. The performance of the simple algorithms that have been used as standard benchmark for comparison is denoted with 'o'. In particular, Figure 1 presents the values of MAE for the CWT variations, where it is also shown how MAE changes as the value of CWT is varied. From this diagram, it can be noted that most algorithm variations demonstrate a MAE between '0.5' and '0.8' (although there are some with significantly higher error values). This is an indication that CWT variations seem to be having a rather stable behavior on the data set of the studied online community. Similarly, Figure 2 illustrates the values for the MNN variations, also showing how they change as the number of neighbors is varied. In the case of the MNN variations of the algorithm, MAE seems to be fluctuating in a higher degree, ranging between '0.4' and '1'. This is an indication that the accuracy of the MNN variations for the particular data set depends on their exact parameterization.

From these diagrams, it appears that several variations (belonging both to CWT and MNN ones) seem to be performing rather satisfactorily on the examined data set in terms of accuracy. To select the appropriate variation for the CELEBRATE community context, we narrowed down our selection to the top-10 ones (in terms of accuracy) that also provide coverage equal or greater than 60%. For these 10 variations, we also examined their execution time. Table 1 demonstrates these 10 variations have performed upon the studied data set.

Based on these results, and keeping in mind that execution time is very important for the online context of the EUN Web portal, we have chosen an algorithm variation that engages the CWT method for the selection of neighborhood, using the Cosine metric for the calculation of similarity between user preferences (with CWT=0.55).

This algorithm variation, ranked 7th in the table 1, offers a combination of rather good accuracy (prediction with MAE of about 0.676 on the scale of '1' to '5'), high coverage (producing a prediction for about 69% of the resources), and rather fast execution (calculating the prediction in about 17 seconds) for the studied data set.

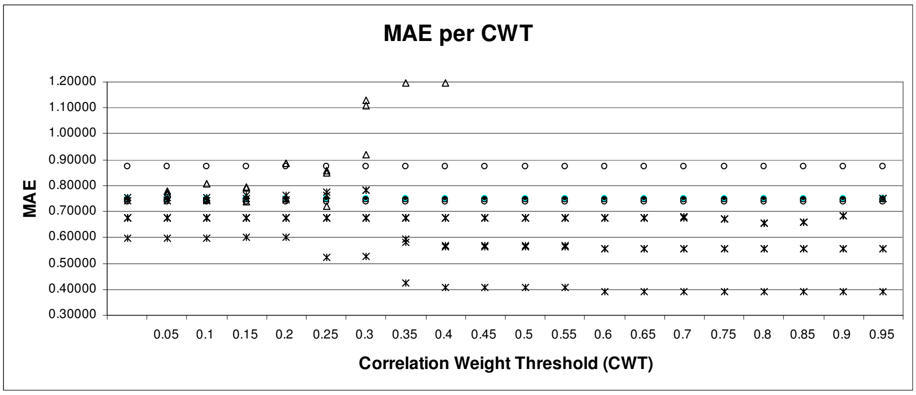

Fig. 1. MAE for each CWT variation.

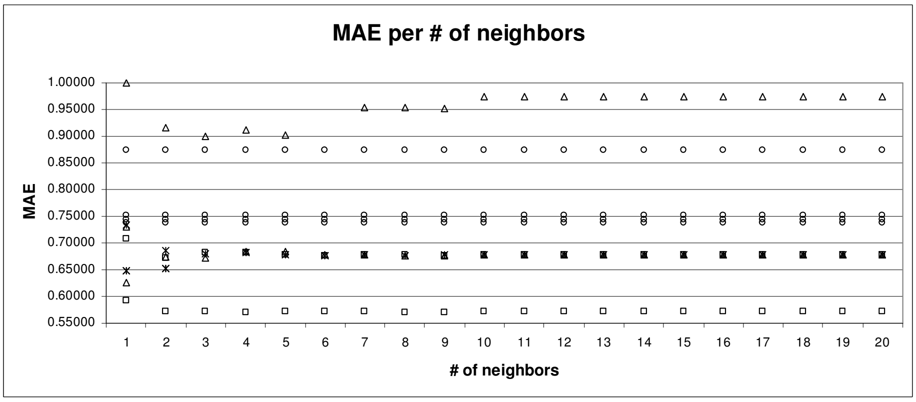

Fig. 2. MAE for each MNN variation.

Table 1. Top-10 algorithm variations according to MAE and execution time (with coverage >65%).

Rank	Variation	Neighb Method	Normalization	MAE	Coverage	Execution Time
1st	Cosine	MNN=4	Deviation-from-Mean	0.57028	69.08%	27 sec.
2nd	Cosine	MNN= 1	Deviation-from-Mean	0.59111	69.08%	23 sec.
3rd	Cosine	CWT=0.85	Simple Mean	0.65388	63.80%	17 sec.
4th	Cosine	CWT=0.85	Weighted Mean	0.65390	63.80%	17 sec.
5th	Euclidian	MNN= 3	Simple Mean	0.67257	69.08%	19 sec.
6th	Pearson	MNN=6	Simple Mean	0.67553	69.08%	22 sec.
7th	*Cosine*	*CWT=0.55*	*Weighted Mean*	*0.67650*	*69.08%*	*17 sec.*
8th	Euclidian	MNN=9	Simple Mean	0.67682	69.08%	19 sec.
9th	Pearson	MNN=8	Simple Mean	0.67685	69.08%	21 sec.
10th	Euclidian	MNN= 14	Simple Mean	0.67718	69.08%	18 sec.

6 Conclusions

This paper experimentally investigated which variation of a particular multi-attribute utility algorithm is more appropriate for an implementation in the collaborative filtering service of the CELEBRATE portal. CollaFiS, a simulation environment for collaborative filtering algorithms, has been used to analyze a collected data set with actual teachers' evaluations and to test various parameterizations of the proposed algorithm. This method made it possible to choose a particular parameterization that seems most appropriate for the online community of this portal.

As shown in other experiments testing the particular algorithm in different application settings, the parameterization of the algorithm greatly depends on the properties of the data set. Therefore results cannot be generalized for other domains. On the other hand, simulation environments such as CollaFiS [8] can help designers in testing their candidate algorithms before their application in a specific case study.

In domains like movie recommendation, the existence of multi-attribute ratings is rare, due to the nature of the domain. On the other hand, in domains like TEL, learning repositories often collect evaluations from experts or users upon multiple criteria, which serve local needs that user should take into consideration when selecting an appropriate resource (e.g. pedagogical quality, technical quality, ease of use [10]. Apart from the CELEBRATE portal, another characteristic case study where there are plenty of multi-attribute evaluations on learning resources is the MERLOT portal (http://www.merlot.org). Therefore, the study of multi-attribute recommendation algorithms is particularly suitable for such application contexts, considering the existing amount of multi-attribute rating data.

It is important to note that the presented simulation experiment may serve as only a first step towards the understanding and appropriate specialization for a collaborative filtering service for the CELEBRATE user community. It has to be further complemented with experiments that will study the needs and expectations of the users, their information seeking tasks, and how recommended resources may be used in the context of their teaching activities. The learning resources recommendation domain cannot be viewed in a way similar to other domains (such as movies recommendation), since the way recommendations are produced and presented can depend on the pedagogical use of the recommender system (e.g. adopting a collaborative problem solving approach). As a consequence, the design principles and pedagogical conditions for such a recommender system constitute a complex problem space where many stakeholders and their needs are to be covered. Although technology is an enabler of the process, the user should stand in the center with his educational specific needs and information seeking tasks at hand. This understanding may prove more important in the context of learning resources recommendation than plain performance testing of competitive algorithms. Nevertheless, we aim to experimentally test the performance of the selected variation also against other multi-attribute algorithms that are proposed in the literature [1].

7 Acknowledgement

Part of the work presented in this paper has been funded by the European Commission, and more specifically the project No EL/06/B/F/PP-148238 “Rural e-Gov: Training SMEs of Rural Areas in using e-Government Services” of the Leonardo da Vinci (LdV) Programme.

References

1. Adomavicius G., Kwon Y., "New Recommendation Techniques for Multi-Criteria Rating Systems", IEEE Intelligent Systems, Special Issue on Recommender Systems, 20(3), 2007.
2. Adomavicius G., Tuzhilin A. “Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”, IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749, 2005.
3. Drachsler H., Hummel H.G.K., Koper R., “Recommendations for learners are different: Applying memory-based recommender system techniques to lifelong learning”, in Workshop on Social Information Retrieval for Technology-Enhanced Learning (SIRTEL’07), Crete, Greece,September 17-20, 2007.
4. Herlocker J.L., Konstan J.A., Terveen L.G., Riedl J.T., "Evaluating Collaborative Filtering Recommender Systems", ACM Transactions on Information Systems, 22(1), 5-53, 2004.
5. Lemire D., Boley H., McGrath S., Ball M., “Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation”, International Journal of Interactive Technology & Smart Education, 2 (3), 2005.
6. Keeney R.L., Value-focused Thinking: A Path to Creative Decisionmaking. Cambridge MA: Harvard University Press, 1992.
7. Manouselis N., Costopoulou C., "Analysis and Classification of Multi-Criteria Recommender Systems", to appear in World Wide Web: Internet and Web Information Systems, 2007.
8. Manouselis N., Costopoulou C., “Designing a Web-based Testing Tool for Multi-Criteria Recommender Systems”, Engineering Letters, Special Issue on “Web Engineering”, 13(3), 2006 (http://www.engineeringletters.com/issues_v13/issue_3/EL_13_3_22.pdf).
9. Manouselis N., Costopoulou C., “Experimental Analysis of Design Choices in a Multi-Attribute Utility Collaborative Filtering System”, International Journal of Pattern Recognition and Artificial Intelligence, Special Issue on "Personalization Techniques for Recommender Systems and Intelligent User Interfaces", 21(2), 311-331, 2007.
10. Nesbit J. C., Belfer K., Leacock T., “Learning object review instrument (LORI)”, E-Learning Research and Assessment Network, 2003 (http://www.elera.net/eLera/Home/About%20%20LORI/).
11. Rafaeli S., Barak M., Dan-Gur Y., Toch E., “QSIA – a Web-based environment for learning, assessing and knowledge sharing in communities”, Computers & Education, 43, 273-289, 2004.
12. Recker, M., Walker, A., Lawless, K., “What do you recommend? Implementation and analyses of collaborative filtering of Web resources for education”, Instructional Science, 31(4/5), 229-316, 2003.
13. Tso K.H.L., Schmidt-Thieme L., "Evaluation of Attribute-aware Recommender System Algorithms on Data with Varying Characteristics", in Ng W.K., Kitsuregawa M., Li J., & Chang K. (Eds.), Advances in Knowledge Discovery and Data Mining, 10th Pacific-Asia Conference, PAKDD 2006, Singapore, April 9-12, 2006.