Revisiting the Multi-Criteria Recommender System of a Learning Portal

Nikos Manouselis[1], Giorgos Kyrgiazos[2], Giannis Stoitsis[1]

[1]Agro-Know Technologies
17 Grammou Str., 152 35, Athens, Greece
{nikosm, stoitsis}@ieee.org

[2]Computer Technology Institute and Press - Diophantus
10 Davaki Srt., 115 26, Athens, Greece
kyrgiazos@cti.gr

Abstract. Results of previous studies have indicated that the same recommendation algorithms perform in totally different ways when a different dataset is considered, thus leading to the need for continuous monitoring of how algorithms perform in a realistic and evolving setting. In this paper we investigate such a real life implementation of a multi-criteria recommender system and try to identify adjustments that need to take place in order for it to better match the requirements of its operational environment. More specifically, we examine the case of a multi-attribute collaborative filtering algorithm that has been supporting the recommendation service within a Web portal for organic and sustainable education.

Keywords: Multi-criteria recommender system, experimental investigation, real life deployment, learning portal

1 Introduction

In the domain of education, recommender systems have been introduced more than a decade ago, with deployed and well-studied systems like Altered Vista (Recker et al., 2003) and CoFIND (Dron et al., 2000). Additionally, surveys of the systems that have been actually implemented in a real life setting indicate that they are very few (Manouselis et al., 2011; Manouselis et al., 2012; Verbert et al., 2012). For instance, a very recent analysis of existing educational recommenders that has been carried out by Manouselis et al. (2012) revealed that out of the forty two (42) systems proposed in the literature since 2000, only thirteen (13) have been actually deployed as a fully implemented and operational system – and not all in a real usage setting. This significantly inhibits the quality of research that can take place in this application domain, since it is important to be able to study the social and psychological requirements on how people react to and act upon recommender systems for the learning sciences (Buder & Schwind, 2012). It also implies that the deployment of a real world recommender system in education is a demanding and challenging exercise.

In this paper, we try to reflect on one of the main questions that the people responsible for an operational recommender system need to face: how can we monitor, test, and fine-tune the algorithms deployed in a real setting, by using data from its actual operation. More specifically, we focus on the case of an existing educational recommender system that collects data that educators and learners provide on digital content that may be used to support education and research on organic and sustainable agriculture, and uses this dataset to provide recommendations about relevant resources (Manouselis et al., 2009). Our study particularly focuses on the collaborative filtering algorithm that has been chosen and parameterized to collect multi-criteria ratings on the content items in order to recommend new items to the users, and tries to investigate two dimensions:

- How does the implemented algorithm perform over a current rating data set from the targeted educational application, also compared to some alternative multi-criteria recommendation algorithms;
- How do the studied algorithms perform over a synthetic data set that simulates how the users of the targeted application will have rated the available content items in a future time instance.

The remainder of this paper is structured as it follows. Section 2 provides the background of this study as it introduces collaborative filtering using multi-criteria ratings, as well as the specific educational application that serves as a case study. Then, Section 3 presents the methodology of this study, by describing the experimental setting and environment in which the study took place, the multi-criteria recommendation algorithms compared, the metrics used for their comparison, as well as the multi-criteria rating data sets that have served as the comparison basis. Section 4 presents the results of our experimental investigation, particularly presenting how the studied algorithms performed over each data set. A discussion of the results, their implication on the implemented real world service, as well as the limitations of this study is included in Section 5. Finally, some overall conclusions and directions of future research are given.

2 Background

2.1 Multi-Criteria Collaborative Filtering

In most recommender systems, the utility function usually considers a single-criterion value, e.g., an overall evaluation or rating of an item by a user. In recent work, this assumption has been considered as limited (Adomavicius & Kwon, 2007; Adomavicius et al., 2011), because the suitability of the recommended item for a particular user may depend on more than one utility-related aspect that the user takes into consideration when making the choice. Particularly in systems where recommendations are based on the opinion of others, the incorporation of multiple criteria that can affect the users' opinions may lead to more accurate recommendations. Thus, the additional information provided by multiple dimensions or criteria could help to improve the quality of recommendations because it makes it possible to represent more complex preferences of each user. Recommender systems have already adopted multiple criteria as relevant research indicates (Adomavicius et al., 2011; Lakiotaki et al., 2011). A recent survey by Adomavicius et al. (2011) identified more than fifty (50) such systems that can be broadly classified as multi-criteria recommender ones.

Multi-criteria collaborative filtering is an extension of traditional collaborative filtering systems that is based on ratings expressed over multiple dimensions describing an item (Adomavicius et al., 2011). They allow a user to specify his individual preferences by rating each item upon multiple criteria, and then recommend to the user the items that can best reflect the user's individual preferences based on the multi-criteria ratings provided by this and other users. In single-attribute (or single-criterion) 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. The recommender estimates a utility function R for the entire or some part of the Users Items space based on known ratings and possibly other information (such as user profiles and/or item features). Collaborative filtering aims to predict this utility R of items for a particular user (called active user) based on the items previously evaluated by other users. That is, the utility $R(a,i)$ of item i for the active user $a \in Users$ is estimated based on the utilities $R(u,i)$ assigned to item i by those users $u \in Users$ who are "similar" to user a.

The difference to single-criterion rating systems is that the utility function $R(u,i)$ is the total utility of an item, calculated by synthesizing the partial utilities of the item on each one of the rating dimensions (or criteria). Assuming that there is no uncertainty during the decision process, the total utility of an item $i \in Items$ for a user $u \in Users$ is often expressed as an additive value function of the evaluation or ratings $g(u,i)$ that user *u* provides for item *i* over each one of the *k* criteria, such as:

$$R(u,i) = \sum_1^k g(u,i) \qquad (1)$$

Such a linear form of the total utility function is the simplest and most popular form of a utility function. Other forms that could be used include an ideal point model, dependencies and correlations, as well as diminishing utility forms (Price & Messinger, 2005).

The collaborative filtering techniques that use multi-criteria ratings to predict an overall rating and/or individual criteria ratings can be classified by the formation of the utility function into two categories: heuristic-based (sometimes also referred to as memory-based) and model-based techniques (Adomavicius et al., 2011). Heuristic-based techniques compute the utility of each item for a user on the fly based on the observed data of the user and are typically based on a certain heuristic assumption. In contrast, model-based techniques learn a predictive model, typically using statistical or machine-learning methods that can best explain the observed data, and then use the learned model to estimate the utility of unknown items for recommendations.

Different approaches may be also followed by the algorithms developed to support multi-criteria collaborative filtering. For instance, algorithms may (Adomavicius & Kwon, 2007; Manouselis & Costopoulou, 2007):

- try to predict the total utility for an item using the total utility values that the item has for other users;
- try to calculate a separate prediction per each criterion and then use the utility function in order to acquire the predicted total utility.

2.2 Case Study

In this paper, we focus on the particular case of a real life implementation of a multi-criteria recommender system in the context of an educational application. This is the case of the Organic.Edunet Web portal for agricultural and sustainable education (http://www.organic-edunet.eu) that was launched in 2010. Its aim has been to facilitate access, usage and exploitation of digital educational content related to Organic Agriculture (OA) and Agroecology (AE). In order to achieve this aim, it networked existing collections with educational content on relevant topics from various content providers, into a large federation where content resources are described according to standard-complying metadata.

After more than two (2) years of operation, Organic.Edunet seems to be established as a reference source of educators and researchers working on relevant topics. It has already attracted more than 20,000 unique visitors from about 150 countries. About 5,000 visitors have registered into the portal's community, being able to receive regular information updates related to the portal and its content, as well as having access to personalized services such as receiving recommendations about potentially interesting content resources. The recommendation service in Organic.Edunet is supported by two separate algorithms that are using different data as input and are currently running independently (Manouselis et al., 2009): a content-based recommender using tags and textual reviews as input; and a multi-criteria collaborative filtering system that uses as input the ratings that users provide over three criteria: Subject Relevance, Educational Value and Metadata Quality.

This study focuses on the multi-criteria algorithm and the recommendations that it produces. This algorithm was proposed by Manouselis & Costopoulou (2007) as a multi-criteria extension to typical heuristic neighborhood-based algorithms that may be found in the collaborative filtering literature. It follows the generic steps of Herlocker et al. (2002) in order to calculate a prediction:

- *Stage A - Similarity Calculation:* similarity between the examined user (active user) and the rest of the users is calculated using some similarity measure;
- *Stage B - Feature Weighting:* further weight similarity according to the characteristics of each examined user or some heuristic rules;
- *Stage C - Neighborhood Formation/Selection:* select the set of users to be considered for producing the prediction;

- *Stage D - Combining Ratings for Prediction:* normalize the ratings that the users in the neighborhood have provided for the unknown item, and use some method to combine them in order to predict its utility for the active user.

The implemented multi-criteria extension is called the **Similarity per evaluation (PG)** algorithm. It calculates the prediction of the total utility $R(a,i)$ of a target item $i \in Items$, by calculating k predictions of how the active user would evaluate i upon each one of the criteria, and then synthesizes these predictions into a total utility value.

Since the implementation of the recommendation algorithm took place during the design stage of the portal, we based our selection on the experience from a lab testing experiment that took place using an existing data set from another learning portal (Manouselis et al., 2010). Results of the simulated execution of more than 360 variations of the PG algorithm over this data set indicated that it would make sense to implement a version that: uses a Cosine/Vector distance function to measure similarity between users; engages a Correlation Weight Threshold (CWT) to select users that have similarity value of more than 0.5 for the neighborhood; and calculates predicted ratings as a weighted mean of the ratings that the neighbors have given over an unknown item. This variation has shown to achieve a Mean Absolute Error (MAE) over the prediction of less than 0.7 (in a scale 1-5) and a coverage close to 70% of the items.

Nevertheless, the fact that the specific algorithm or variation performed well over a data set coming from a similar application context (that is, of a portal with learning resources) does not mean that it would also perform well during the operation of the Organic.Edunet portal. There are several reasons for this:

- The properties of the users vs. items matrix of Organic.Edunet may be different than the ones of the dataset of the other application.
- The properties of the Organic.Edunet matrix may change/evolve with time.
- Alternative algorithms (e.g. new ones proposed in literature) that were not included in the initial experimentation may prove to perform better than the one selected.

To this end, we decided to repeat the experimental investigation of candidate algorithms for the Organic.Edunet portal, using additional algorithms as options, as well as a synthetic data set that tries to mimic a future state where users will have provided more multi-criteria ratings over the educational resources.

3 Methodology

3.1 Experimental Setting

The main goal of the experimental testing has been to investigate the performance of different variations of both the algorithm currently implemented in Organic.Edunet as well as alternative multi-criteria recommendation algorithms. The specific objectives have been:

- To use a current instance of the users vs. items matrix of Organic.Edunet in order to execute all candidate variations and measure their expected performance.
- To generate a synthetic data set that mimics an instance of the Organic.Edunet community in the future, and explore if performance of the candidate algorithms would be expected to change in the future.

The evaluation protocol follows the typical steps of offline experiments with pre-collected or simulated data that Shani & Gunawardana (2011) also described for testing the performance of candidate algorithms. Generally speaking, our experiment follows the approach of similar experiments in other domains (Herlocker et al., 2004) or education (Lemire et al., 2005; Sicilia et al., 2010). The following paragraphs describe the settings, methods and tools of the experimental investigation.

The offline experiment took place using a software environment that has been specifically developed and used for the simulation of multi-criteria recommender systems, called the ***Collaborative Filtering Simulator (CollaFiS).*** This environment allows for importing various data sets, parameterizing candidate algorithms, executing them and measuring expected performance using multiple performance metrics (Manouselis & Costopoulou, 2006). The ***CollaFiS*** environment has been extended to support the algorithms and metrics that are particularly studied in this experiment, as described later. ***CollaFiS*** provides the option for experimentally testing the multi-criteria algorithms proposed by Manouselis & Costopoulou (2007). We have extended the previous implementation of ***CollaFiS*** in order to also include some algorithms proposed by Adomavicius & Kwon (2007). Overall, the studied algorithms included:

- the ***Similarity per evaluation (PG)*** algorithm (currently implemented in Organic.Edunet) that calculates similarity separately upon each criterion, predicts the rating also separately upon each criterion, and then is synthesizing the predictions into a total predicted utility;
- the ***Average Similarity (AS) and Minimum or Worst-case Similarity (WS)*** algorithm versions proposed by Adomavicius & Kwon (2007) that use either the average or the minimum of the similarities over each criterion in order to calculate the total predicted utility;
- some ***Non-personalised*** algorithms as a basic comparison, such as giving random values as predictions or calculating an arithmetic, geometrical or deviation-from-mean weighted sum of all past evaluations.

For the personalized algorithms ***(PG, AS, WS)*** we have considered the following design options in order to study different variations within the generic steps of Herlocker et al. (2002) described in section 2.2:

- *during Stage A - Similarity Calculation:* examined the calculation of the similarity using the Euclidian, Vector/Cosine, and Pearson distance functions as options.
- *during Stage C - Neighborhood Formation/Selection:* examined both the use of a *Correlation Weight Threshold* (CWT) for the similarity value as a selector of potential neighbors, as well as of an absolute value for the *Maximum Number of Neighbors* (MNN).
- *during Stage D - Combining Ratings for Prediction:* examined three different options for synthesizing partial utilities, i.e. calculating the prediction as a simple arithmetic mean, as a mean weighted by the similarity value, as well as a normalized weighted mean that takes into consideration also the deviation from the arithmetic mean (as Herlocker et al. 2002 suggest).

This led to 18 variations of each examined algorithm. By also experimenting with various values for the CWT (20 variations between '0' and '1' as a threshold) and MNN (20 variations using '1' to '20' maximum neighbors) parameters, the number grew to more than 1,080 algorithmic variations explored in total.

There are several performance metrics used in the literature. In this experiment we examined the following evaluation metrics that ***CollaFiS*** incorporates:

- *Accuracy*: to measure the predictive accuracy of the multi-criteria algorithms, we calculated the mean-absolute error (MAE). MAE is the most frequently used metric when evaluating recommender systems. Herlocker et al. (2004) have demonstrated that since it is strongly correlated with many other proposed metrics for recommender systems, it can be preferred as easier to measure, having also well understood significance measures.
- *Coverage*: to measure the coverage of the multi-criteria algorithms, we calculated the items for which an algorithm could produce a recommendation, as a percentage of the total number of items. Herlocker et al. (2004) recommend the measurement of coverage in combination with accuracy.

Two different data sets have been used to support the simulated execution of the algorithms. Both have been imported into ***CollaFiS*** and appropriately processed. To facilitate the execution of the experiments, they have been split into one training and one testing component (using an 80%–20% split).

The first data set (OEreal) has been a recent export/instance of the users vs. items matrix of Organic.Edunet, with the collected multi-criteria ratings that the users of the portal have provided over the content items. As mentioned before, Organic.Edunet collects user evaluations over three criteria that are all rated using a discrete scale from 1 to 5. In this real dataset, 99 users have provided 477 multi-criteria ratings over 345 items.

The second data set was a simulated one (OEsim) that tried to represent a future state of the *Users x Items* matrix of Organic.Edunet. More specifically, the distributions of the ratings of the OAreal dataset were taken as input to a Monte Carlo generator of random multi-criteria ratings of the same users. The considered scenario is that the current users that have been rating a sample of the Organic.Edunet items provide more ratings on this specific sample of already rated items in order to make it more dense. The produced synthetic dataset incorporates the original real one, has the same number of users and items, and includes a total number of 1,280 multi-criteria ratings. In a similar way, alternative scenarios could be considered, with more users and/or items, and with more dense or sparse data sets.

Table 1. Non-personalized algorithms over OEreal and OEsim data sets

Variation	**Pure Random**	**Random Existing Rating**	**Arithmetic Mean**	**Geometric Mean**	**Deviation-from-Mean**
MAE over OEreal	1.59	1.33	1.28	1.27	1.03
Coverage over OEreal	100%	100%	37.89%	37.89%	32.63%
MAE over OEsim	1.38	1.01	0.86	0.89	0.83
Coverage over OEsim	100%	100%	96.09%	96.09%	96.09%

Table 2. Top-5 CWT and top-5 MNN variations over OEreal and OEsim data sets

Algorithm	**Similarity**	**Normalization method**	**AVG Coverage**	**AVG MAE**
Top-5 MNN variations over OEreal dataset				
PG	Cosine	Deviation-from-Mean	18.95%	0.9928
PG	Euclidian	Simple Mean	18.95%	1.3194
PG	Cosine	Weighted Mean	18.95%	1.3337
AS	Euclidian	Deviation-from-Mean	18.95%	1.5008
WS	Cosine	Deviation-from-Mean	18.95%	1.6886
Top-5 CWT variations over OEreal dataset				
PG	Cosine	Deviation-from-Mean	16.32%	0.8650
PG	Cosine	Simple Mean	16.32%	1.1831
AS	Cosine	Deviation-from-Mean	16.95%	1.5202
WS	Cosine	Deviation-from-Mean	16.37%	1.7316
AS	Cosine	Simple Mean	16.95%	2.1074

Top-5 MNN variations over OEsim dataset				
PG	Euclidian	Simple Mean	0.6133	0.8626
PG	Cosine	Simple Mean	0.6133	0.8653
PG	Cosine	Deviation-from-Mean	0.6133	0.8855
AS	Euclidian	Deviation-from-Mean	0.6133	1.2972
WS	Euclidian	Deviation-from-Mean	0.6133	1.8486
Top-5 CWT variations over OEsim dataset				
PG	Cosine	Simple Mean	0.5791	0.8673
PG	Cosine	Weighted Mean	0.5791	0.8681
PG	Cosine	Deviation-from-Mean	0.5791	0.8908
AS	Cosine	Deviation-from-Mean	0.5934	1.2983
AS	Cosine	Weighted Mean	0.5934	2.2086

4 Results

In this section we will present the results that have been produced by the CollaFiS tool, after executing all the studied variations of the algorithms. For each dataset we are going to present top-5 sets of options, based on metrics mentioned above.

4.1 Real Data Set

The execution of the candidate algorithms over the OEreal dataset did not provide very good results. It seems that the majority of the tested variations performed a bit better than the non-personalised algorithms (Table 1) but still with a very low coverage that was in the vicinity of 16%-18% of the items for which a prediction needed to be made. As Table 2 shows, there are some variations (like **the PG Cosine Deviation-from-Mean** with both MNN and CWT parameters) that had an acceptable MAE that is below '1'. Still we consider this error to be rather high for an operational recommender system. These results imply that the performance of any algorithm would be judged not satisfactory if only the current data set of Organic.Edunet was used for experimentation.

In Figures 1 and 2 (and in all diagrams of this paper), the continuous lines are used to illustrate the performance of ***PG*** variations, the heavy dashed ones the ***AS*** variations, and the lightly dashed ones the ***WS*** variations. These diagrams illustrate that the ***PG*** variations seem to be generally performing better than the ***AS*** and ***WS*** ones. Nevertheless, this performance seems to be rather low over the OEreal data set.

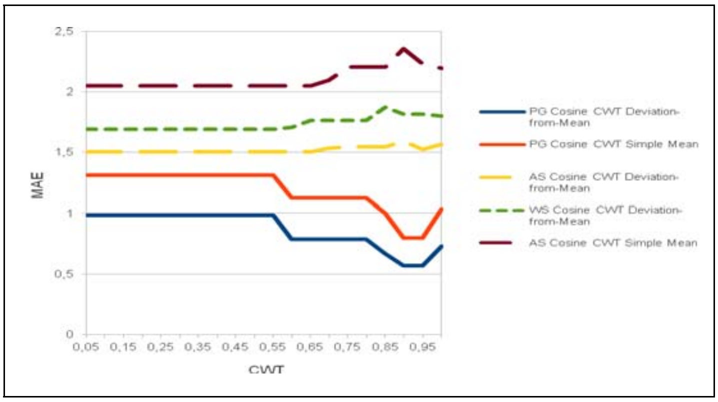

Fig. 1. MAE metric performance for top-5 CWT variations over OEreal data set

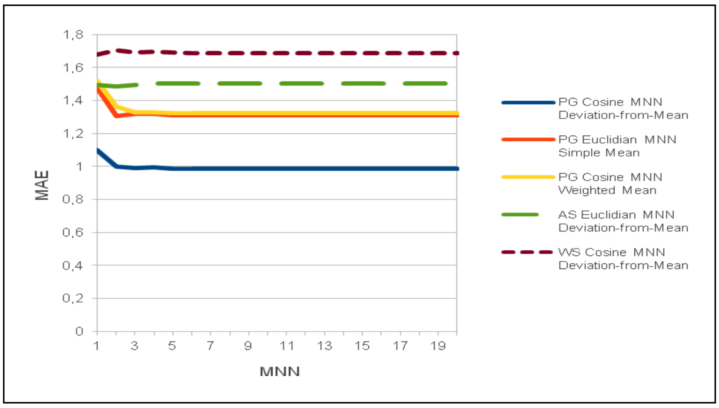

Fig. 2. MAE metric performance values for top-5 MNN variations over OEreal data set

4.2 Synthetic Data Set

The execution of the candidate algorithms over the synthetic OEsim dataset seemed to perform much better over the original OEreal one, as one would have expected (since a more dense version of the dataset has been created). As it is also shown in Table 2, the majority of the outstanding variations have a rather good coverage that is close to (for CWT) and more than (for MNN) 60%.

Surprisingly the MAE results seem to be at the level of the ***non-personalised*** algorithms (also presented in Table 1) and around 0.86 for the ***PG MNN Euclidian Simple Mean*** and the ***PG CWT Cosine Simple Mean***. It seems that very simple algorithms that create weighted sums of the past ratings, such as the ***Arithmetic Mean*** and the ***Geometrical Mean***, may provide predictions that have less MAE than the collaborative filtering variations. This could be due to the fact that the simulated users have provided additional ratings with similar distributions but still such a simplistic interpretation of this observation is not enough. Again, the graphical illustrations of

Figures 3 and 4 show that in most cases the ***PG*** variations seem to be performing better than the ***AS*** and ***WS*** ones, although the differences are small. To further investigate which would be the more appropriate algorithm variations to support recommendation in Organic.Edunet in such a future scenario, we did an additional experimental analysis.

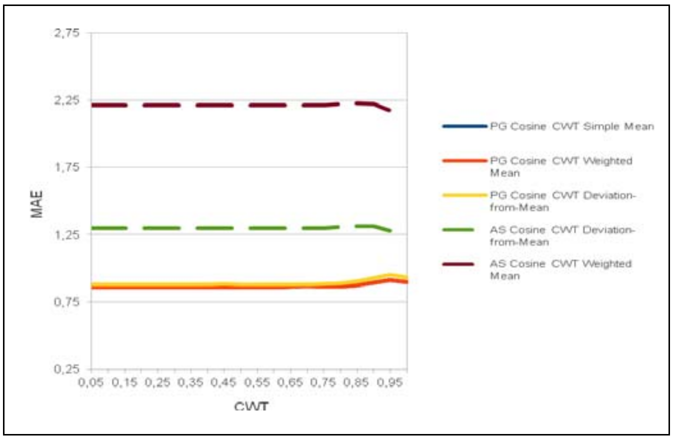

Fig. 3. MAE metric performance for top-5 CWT variations from all candidate variations over OEsim data set

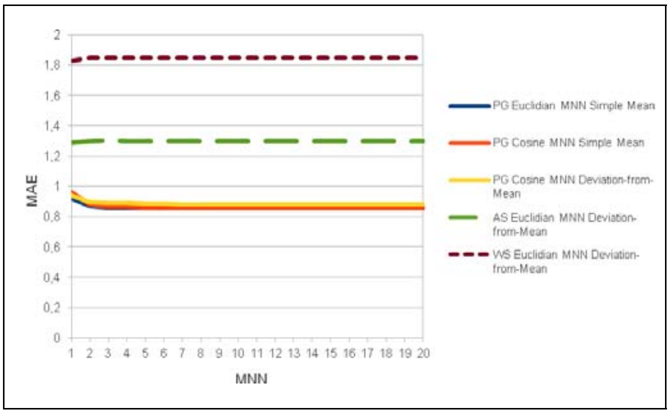

Fig. 4. MAE metric performance for top-5 MNN variations from all candidate variations over OEsim data set

More specifically, we investigated the performance of the two algorithms that performed better over the real dataset (OEreal) also over the synthetic OEsim. In this way we tried to examine if some of the algorithm variations that performed in a good way over current data, would also perform in a similar way over a future state of Organic.Edunet. As illustrated in Table 3, these algorithm variations seemed to also perform in a satisfactory way over the OEsim data. Some of them seem to be common across all data sets, with most prominent being the ***PG Cosine Deviation-from-Mean*** variation.

Table 3. Performance of top-2 CWT and top-2 MNN variations of OEreal data set over the OEsim one

Algorithm	Similarity	Normalization method	AVG Coverage	AVG MAE
MNN variations				
PG	Cosine	Deviation-from-Mean	61.33%	0.8855
PG	Euclidian	Simple Mean	61.33%	0.8626
CWT variations				
PG	Cosine	Deviation-from-Mean	57.91%	0.8908
PG	Cosine	Simple Mean	57.91%	0.8673

The experimental analysis overall indicates that the PG algorithm currently implemented in the Organic.Edunet portal is still a good choice. Yet, its exact parameterization and fine-tuning so that the right values are chosen that will give better results, is an exercise that needs to be taking place quite often in such a changing environment. As the community of users grows, the properties of the Users x Items matrix (that is, of the dataset) will be dynamically changing. For instance, during the past year only, more than 1,000 new users have registered in the portal. In addition, the content collections to which the portal gives access to, is ready to expand from about 11,000 items to some 30,000.

This calls for careful consideration and planning from the perspective of the designer and operator of the recommendation service. One option would be to run frequent offline experiments with most recent updates of the data set, in order to find which algorithm variations is more appropriate every time for the application. Another approach would be the investigation of adaptive algorithms that will automatically measure their performance (e.g. the accuracy and coverage of their predictions) over a dataset with specific properties, and then adapt their parameters in order to achieve better results. Such an approach can be a rather computationally-demanding task that calls for a re-engineering of the existing recommendation service of the portal and maybe an investigation of new multi-criteria recommendation algorithms.

5. Conclusions

In this paper we investigated how the recommendation algorithm used in a real life implementation of a multi-criteria recommender system performs under various experimental conditions, by using as input different datasets with multi-criteria ratings. The case study has been a portal for organic and sustainable education, and the experimentation tested a wide number of variations with one real dataset and one synthetic one. The results indicated that some particular variations seem to perform in a satisfactory way over both datasets. This was an interesting observation, considering that in related work we have witnessed significant alterations in the performance of the same algorithms over different datasets (Manouselis & Costopoulou 2007;

Manouselis et al., 2010). It gave useful input regarding the improvements that need to be made in the algorithm currently implemented in Organic.Edunet.

Our future work includes a more extensive experiment where the correlation between the various algorithmic parameters and options and the properties of the data sets will be explored. The currently available real data sets will be used as generators of a large number of synthetic data sets with varying properties. Then the ***CollaFiS*** simulator will be used to execute a large number of variations and measure how they perform over the various data sets. Additional metrics will also be engaged, such as typical ones found in information retrieval (e.g. Precision, Recall, F-measure).

6. Acknowledgements

The work presented in this paper has been funded with support by the European Commission, and more specifically the FP7 project "agINFRA: A data infrastructure to support agricultural scientific communities" of the Research Infrastructures, Capacities Programme, Objective INFRA 2011 1.2.2: Data infrastructures for e-Science. Giorgos Kyrgiazos was affiliated with (and funded by) Agro-Know Technologies at the time that this study took place.

References

1. Adomavicius, G., Kwon, Y.: New Recommendation Techniques for Multi-Criteria Rating Systems. In: IEEE Intelligent Systems, 22(3), 48-55 (May/June 2007) .
2. Adomavicius, G., Manouselis, N., Kwon Y.: Multi-Criteria Recommender Systems. In Kantor, P., Ricci, F., Rokach, L., Shapira, B. (Eds.), Recommender Systems Handbook: A Complete Guide for Research Scientists & Practitioners, Springer, 769-803 (2011)
3. Buder, J., Schwind, C.: Learning with personalized recommender systems: A psychological view. In: Computers in Human Behavior, 28, 207-216 (2012)
4. Dron, J., Mitchell, R., Siviter, P., Boyne, C.: CoFIND-an experiment in n-dimensional collaborative filtering. In: Journal of Network and Computer Applications, 23(2), 131-142 (2000)
5. Herlocker, J., Konstan, J., Riedl, J.: An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. In: Information Retrieval, 5, 287–310 (2002)
6. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating Collaborative Filtering Recommender Systems. In: ACM Transactions on Information Systems, 22(1), 553 (2004)
7. Lakiotaki, K., Matsatsinis, N.F., Tsoukiàs, A.: Multi-Criteria User Modeling in Recommender Systems. In: IEEE Intelligent Systems, 26(2), 64-76 (2011)
8. Lemire, D., Boley, H., McGrath, S., Ball, M.: Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation. In: International Journal of Interactive Technology and Smart Education, 2(3) (2005)
9. Manouselis, N., Costopoulou, C.: Designing a Web-based Testing Tool for Multi-Criteria Recommender Systems. In: Engineering Letters, Special Issue on Web Engineering, 13(3), http://www.engineeringletters.com/issues_v13/issue_3/EL_13_3_22.pdf (2006)

10. Manouselis, N., Costopoulou, C.: Experimental Analysis of Design Choices in Multi-Attribute Utility Collaborative Filtering. In: International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 21(2), 311-331 (2007)
11. Manouselis, N., Drachsler, H., Verbert, K., Duval, E.: Recommender Systems for Learning. In: Springer Briefs (2012)
12. Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., Koper, R.: Recommender Systems in Technology Enhanced Learning. In: Kantor, P, Ricci, F, Rokach, L, Shapira, B (eds), Recommender Systems Handbook, pp. 387-415, Springer US (2011)
13. Manouselis, N., Kosmopoulos, T., Kastrantas, K.: Developing a Recommendation Web Service for a Federation of Learning Repositories. In: Proc. of the International Conference on Intelligent Networking and Collaborative Systems (INCoS), IEEE Press (2009)
14. Manouselis, N., Vuorikari, R., Van Assche, F.: Collaborative Recommendation of e-Learning Resources: An Experimental Investigation. In: Journal of Computer Assisted Learning, 26(4), 227-242 (2010)
15. Price, B, Messinger, P.R.: Optimal Recommendation Sets: Covering Uncertainity over User Preferences. In: Informs Annual Meeting: Denver 2004, October 24-27 2004, AAAI Press (2005)
16. Recker, M.M., Walker, A., Lawless, K.: What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. In: Instructional Science, 31(4/5), 299–316 (2003)
17. Shani, G., Gunawardana, A. Evaluating Recommendation Systems. In Kantor, P, Ricci, F, Rokach, L, Shapira, B (eds), Recommender Systems Handbook, 257-297, Springer US, (2011)
18. Sicilia, M.-A., Garca-Barriocanal, E., Sanchez-Alonso, S., Cechinel, C.: Exploring user-based recommender results in large learning object repositories: the case of MERLOT. In Manouselis, N, Drachsler, H, Verbert, K, Santos, OC (eds) Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010), Procedia Computer Science, 1(2):2859-2864 (2010)
19. Verbert, K., Manouselis, N., Xavier, O., Wolpers, M., Drachsler, H., Bosnic, I., Duval, E., Context-aware recommender Systems for Learning: a Survey and Future Challenges, IEEE Transactions on Learning Technologies, https://lirias.kuleuven.be/handle/123456789/338644 (2012)