Collaborative recommendation of e-learning resources: an experimental investigation

N. Manouselis,* R. Vuorikari† & F. Van Assche†
*Greek Research & Technology Network (GRNET S.A.), 56 Mesogeion Av., 115 27 Athens, Greece
†European Schoolnet (EUN), 61 Rue de Treves, Brussels, 1040 Brussels, Belgium

Abstract
Repositories with educational resources can support the formation of online learning communities by providing a platform for collaboration. Users (e.g. teachers, tutors and learners) access repositories, search for interesting resources to access and use, and in many cases, also exchange experiences and opinions. A particular class of online services that take advantage of the collected knowledge and experience of users are collaborative filtering ones. The successful operation of such services in the context of real-life applications requires careful testing and parameterization before their actual deployment. In this paper, the case of developing a learning resources’ collaborative filtering service for an online community of teachers in Europe was examined. More specifically, a data set of evaluations of learning resources was collected from the teachers that use the European Schoolnet’s learning resource portal. These evaluations were then used to support the experimental investigation of design choices for an online collaborative filtering service for the portal’s learning resources. A candidate multi-attribute utility collaborative filtering algorithm was appropriately parameterized and tested for this purpose. Results indicated that the development of such systems should be taking place considering the particularities of the actual communities that are to be served.

Keywords
collaborative filtering, e-learning, recommender systems.

Introduction
During the last 20 years, recommender systems have gained importance, leading to an abundance of real-life applications on the Web. Such systems assist information retrieval (IR) tasks by providing users with personalized recommendations about online content and services (Miller et al. 2004). In the field of computer-assisted learning, 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 (ADL 2003; Tzikopoulos et al. 2007; Ochoa & Duval 2009). Thus, these users would probably benefit from online guidance and services that will help them identify suitable learning resources from a potentially overwhelming variety of choices. As a consequence, a number of recommender systems that aim to support users find learning resources online have already been introduced (e.g. Anderson et al. 2003; Recker & Walker 2003; Recker et al. 2003; Walker et al. 2004; Lemire et al. 2005; Rafaeli et al. 2005).

A class of recommender systems that attracts high research and commercial interest is collaborative filtering ones. A collaborative filtering system recommends a user with items that people with similar preferences liked in the past. A vast number of studies exist that deals with the design, development and evaluation of collaborative filtering systems (Adomavicius & Tuzhilin 2005). Nevertheless, related research identifies that a common drawback of such systems is that, often, they do not consider the needs of the actual online...
community that they aim to support (Herlocker et al. 2004). It is common to witness evaluation studies of proposed systems that engage usage data from totally different application domains, overlooking the particularities of their own domain (Tso & Schmidt-Thieme 2006). 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 before their actual deployment in real settings, has been outlined (Herlocker et al. 2004; Manouselis & Costopoulou 2007a).

In this direction, the present paper describes an experimental investigation of a collaborative filtering service for an existing web portal based on authentic data that come from the online community that actually uses the portal. The case under study is the CELERATE project led by European Schoolnet (EUN), and the user community is composed of teachers around Europe that access and use 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. In this study, the multi-attribute evaluations serve as input for the collaborative filtering service, which recommends resources that other members of the community found useful. More specifically, the aim of this study is to experimentally investigate which variation of a particular algorithm can be implemented to support the collaborative filtering service of the learning resource portal, which is currently known as Learning Resource Exchange (http://lreforschools.eun.org, hereafter referred as the portal). For this reason, a simulation environment for collaborative filtering algorithms are used to analyse the authentic teachers’ evaluations, to test various parameterizations of the candidate algorithm and to reach to a conclusion about which parameterization is the most appropriate for this online community.

The structure of the paper is the following. First, we review related work by introducing collaborative filtering, study various applications for the recommendation of learning resources and assess their status of development and evaluation. Next, the targeted user community that shares learning resources through the portal is described. Then, the experimental investigation of the proposed multi-attribute algorithm for collaborative filtering service is carried out. The experimental setting in which the simulated study of the proposed algorithm took place is introduced, and the results of a simulation experiment are illustrated. A discussion of the implications of this study on the enrichment of the portal’s services is outlined including the design dimensions of recommender systems for online learning communities. Finally, the conclusions of this study and directions for future work are outlined.

Background and related work

Collaborative filtering

Internet users are often overwhelmed by the flow of online information, hence the need for adequate systems that will help them manage such situations (Hanani et al. 2001). Recommender systems attempt to guide the user in a personalized way to interesting and useful items in a large space of possible options by producing individualized recommendations as output (Burke 2002). In the mid-1980s, Malone et al. (1987) provided a fundamental categorization of systems that support access to highly dynamic information resources (Belkin & Croft 1992; Baudisch 2001; Hanani et al. 2001). More specifically, they distinguished cognitive filtering systems as the ones that characterize the contents of an information resource (shortly referred to as an item) as well as the information needs of potential item users, and then use these representations to intelligently match items to users, and sociological filtering systems as the ones that are working based on the personal and organizational interrelationships of individuals in a community. Collaborative filtering systems are the most prominent representatives of the second category.

In general, the problem of collaborative filtering is to predict how well a user will like an item that he or she 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 (Herlocker et al. 2002; Adomavicius & Tuzhilin 2005). 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). Table 1 presents an example of how a user-rating matrix looks similar to, and which ratings could be used for the prediction of the...
rating on some particular item that a targeted user has not previously viewed (‘?’).

The collaborative filtering problem may be mathematically formulated as it follows (Adomavicius & Tuzhilin 2005; Manouselis & Costopoulou 2007a, 2007b): 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 $U^c(s): C \times S \rightarrow \mathbb{R}^*$ and measures the appropriateness of recommending an item $s$ to user $c$. It is assumed that this function is not known for the whole $C \times S$ 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 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 (Adomavicius & Tuzhilin 2005). 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 $a$.

When more than one item attribute is used, then the recommendation is formulated as a multi-attribute problem (Adomavicius et al. 2010). Table 2 presents an example of how a user-rating matrix looks similar to in the case of multi-attribute collaborative filtering. In this case, the whole set of ratings upon the attributes of an item can be used for the prediction of either the overall utility of an unknown item or the rating of some particular attribute.

Several algorithmic approaches have been applied to support prediction of unknown ratings in collaborative filtering. In general, they can be distinguished to memory-based (or heuristic-based) approaches and to model-based approaches (Breese et al. 1998; Adomavicius & Tuzhilin 2005). Memory-based algorithms operate as heuristics that make rating predictions based on the entire collection of previously rated items by the users. In contrast, model-based algorithms use the collection of ratings in order to build a model, which is then used to make rating predictions. There are also some hybrid approaches, combining these two types of collaborative filtering algorithms (Xue et al. 2005).

| Table 1. An example user-rating matrix (with ‘?’ the rating that will be predicted). |
|-----------------------------|----------------|---------------|---------------|---------------|---------------|
| Matrix | Life | Star Wars | E.T. | Avatar |
| Riina | 5 | 7 | 5 | 7 | ? |
| Nikos | 5 | 7 | 5 | 7 | 9 |
| Frans | 6 | 6 | 6 | 6 | 5 |

The overall utility of an item for a user in bold, whereas with regular font, the ratings upon three attributes, e.g. story, actors, visuals (with ‘?’ the multi-attribute rating that needs to be predicted).

| Table 2. An example multi-attribute user-rating matrix. |
|-----------------------------|----------------|---------------|---------------|---------------|---------------|
| Matrix | Life | Star Wars | E.T. | Avatar |
| Riina | 5, 4, 7, 5 | 7, 4, 4, 8 | 5, 2, 2, 6 | 7, 4, 4, 7 | ? , ? , ? |
| Nikos | 5, 2, 5, 6 | 7, 8, 4, 4 | 5, 7, 9, 2 | 7, 9, 8, 6 | 9, 6, 7, 5 |
| Frans | 6, 4, 5, 9 | 6, 4, 4, 7 | 6, 1, 2, 7 | 6, 3, 3, 5 | 5, 2, 5, 7 |

In the domain of technology-enhanced learning, a number of recommender systems have been introduced in order to propose learning resources to users. Such systems could potentially play an important educational role, considering the variety of learning resources that are published online (Tzikopoulos et al. 2007) and the benefits of collaboration between tutors and learners (Recker & Wiley 2000, 2001; Nesbit et al. 2002; Kumar et al. 2005). Initial hints of relating collaborative filtering to education have appeared in early relevant papers (Terveen et al. 1997; Chislenko 1998). In this section, we review related literature on representative recommender systems for learning resources.

One of the first attempts to develop a collaborative filtering system for learning resources has been the Altered Vista system (Recker & Walker 2003; Recker et al. 2003; Walker et al. 2004). The aim of this study was to explore how to collect user-provided evaluations of learning resources and then to propagate them in the form of word-of-mouth recommendations about the qualities of the resources. The team working on Altered Vista explored several relevant issues, such as the design of its interface (Recker & Wiley 2000), the development of non-authoritative metadata to store user-provided evaluations (Recker & Wiley 2001), the design...
of the system and the review scheme it uses (Recker & Walker 2003) as well as results from pilot and empirical studies from using the system to recommend to the members of a community both interesting resources and people with similar tastes and beliefs (Recker et al. 2003; Walker et al. 2004).

Another system that has been proposed for the recommendation of learning resources is the Rule-Applying Collaborative Filtering (RACOFI) Composer system (Anderson et al. 2003; Lemire 2005; Lemire et al. 2005). RACOFI combines two recommendation approaches by integrating a collaborative filtering engine that works with ratings that users provide for learning resources with an inference rule engine that is mining association rules between the learning resources and using them for recommendation. RACOFI studies have not yet assessed the pedagogical value of the recommender, nor have they reported some evaluation of the system by users. The RACOFI technology is supporting the commercial site inDiscover (http://www.indiscover.net) for music tracks recommendation. In addition, other researchers have reported adopting RACOFI’s approach in their own systems as well (Fiadhi 2004).

The Questions Sharing and Interactive Assignments (QSIA) for learning resources sharing, assessing and recommendation has been developed by Rafaeli et al. (2004, 2005). This system is used in the context of online communities in order to harness the social perspective in learning and to promote collaboration, online recommendation and further formation of learner communities. Instead of developing a typical automated recommender system, Rafaeli et al. chose to base QSIA on a mostly user-controlled recommendation process. That is, the user can decide whether to assume control on who advises (friends) or to use a collaborative filtering service. The system has been implemented and used in the context of several learning situations, such as knowledge sharing among faculty and teaching assistants, high school teachers and among students, but no evaluation results have been reported so far (Rafaeli et al. 2004, 2005).

In this strand of systems for collaborative filtering of learning resources, the CYCLADES system (Avancini & Straccia 2005) has proposed an environment where users search, access and evaluate (rate) digital resources available in repositories found through the Open Archives Initiative (OAI, http://www.openarchives.org). Informally, OAI is an agreement between several digital archives providers in order to offer some minimum level of interoperability between them. Thus, such a system can offer recommendations over resources that are stored in different archives and accessed through an open scheme. The recommendations offered by CYCLADES have been evaluated through a pilot study with about 60 users, which focused on testing the performance (predictive accuracy) of several collaborative filtering algorithms.

A different approach to learning resources’ recommendation has been followed by Shen and Shen (2004). They have developed a recommender system for learning objects that is based on sequencing rules that help users be guided through the concepts of an ontology of topics. The rules are fired when gaps in the competencies of the learners are identified, and then appropriate resources are proposed to the learners. A pilot study with the students of a Network Education college has taken place, providing feedback regarding the users’ opinion about the system.

Tang and McCalla (2003, 2004a,b,c, 2005) proposed an evolving e-learning system, open into new learning resources that may be found online, which includes a hybrid recommendation service. Their system is mainly used for storing and sharing research papers and glossary terms among university students and industry practitioners. Resources are described (tagged) according to their content and technical aspects, but learners also provide feedback about them in the form of ratings. Recommendation takes place both by engaging a Clustering Module (using data clustering techniques to group learners with similar interests) and a Collaborative Filtering Module (using classic collaborative filtering techniques to identify learners with similar interests in each cluster). The authors studied several techniques to enhance the performance of their system, such as the usage of artificial (simulated) learners (Tang & McCalla 2004c). They have also performed an evaluation study of the system with real learners (Tang & McCalla 2005).

Another system that adopts a hybrid approach for recommending learning resources is the one recently proposed by Drachsler et al. (2007, 2009). The authors build upon a previous simulation study by Koper (2005) in order to propose a system that combines social-based (using data from other learners) with information-based (using metadata from learner profiles and learning activities) recommendation techniques. A further
evaluation of their system has been presented in Nadolski et al. (2009). In this paper, their simulation environment for different combination of recommendation algorithms in a hybrid recommender system has been used in order to compare algorithms regarding their impact on learners in informal learning networks. The same approach is followed by the proposed Learning Object Recommendation Model (LORM) that also follows a hybrid recommendation algorithmic approach and that describes resources upon multiple attributes but has not yet reported to be implemented in an actual system (Tsai et al. 2006).

The benefits of deploying recommender systems for learning resources have been discussed in several recent conceptual studies. For example, there are two interesting papers proposing the use of a recommender to enhance the discovery of learning resources. Downes (2004) suggested that multidimensional evaluations may better reflect differences in the various quality aspects of learning resources and should be stored in user-provided metadata. The author proposed to use the information stored in such metadata in order to recommend resources to users based both on their content-related information, as well as on multi-attribute ratings. Manouselis et al. (2007) tested the multi-attribute rating approach for collaborative filtering of learning objects. Duval (2006) outlined another important aspect of discovering learning approaches online: minimizing users’ efforts. He considered a Google-like approach where something similar to Google’s PageRank technology will determine how useful people have found learning resources, without explicitly asking them how they have liked the resource, but rather by determining certain variables that could be detected from different sources (e.g. tracking users actions with the resource as discussed by Najjar et al. 2006).

Collaborative filtering often shares common ground with social navigation (Munro et al. 1999; Farzan & Brusilovsky 2006, 2008), which might be also seen as a social recommender approach. CoFIND by Dron et al. (2000a,b) is an example of such a system that has been used in education. CoFIND extracted the user model and turned it inside out, creating dimensions of pedagogical metadata used for rating. Furthermore, Brusilovsky et al. (2005) used group filtering in the Knowledge Sea, demonstrating a related use of social processes for recommendation in an open corpus environment with some collaborative filtering characteristics. It might be also worthwhile to mention related work in group filtering where recommendations are made for groups rather than individuals, which can be important in traditional institutional learning communities such as classes (e.g. Masthoff 2005).

Finally, during the special session on ‘Social Software for Professional Learning: Examples and Research Issues’ of the 6th IEEE International Conference on Advanced Learning Technologies (Klamma et al. 2006), the role of recommender systems to enhance the social and collaboration aspects of e-learning systems has been outlined. Such systems may enhance the creation of social networks that will facilitate bottom-up socialization; that is, help people build new relationships and enable them to join learning communities based on their preferences. The introduction of recommender systems (and particularly, collaborative filtering ones) to support such activities has been discussed by other researchers as well (e.g. Dron 2006). Recent contributions in the field include the papers from Koutrika et al. (2009), Gomez-Albarran and Jimenez-Diaz (2009) and Khribi et al. (2008).

Assessing current status

Despite the increasing number of systems proposed for recommending learning resources, a closer look to the current status of their development and evaluation reveals the lack of systematic evaluation studies in the context of real-life applications. Table 3 provides an overview of the selected systems and illustrates several interesting points:

- more than half of the proposed systems (five out of nine) still remain at a design or prototyping stage of development;
- two systems (RACOFI and CYCLADES) have been evaluated by using public data sets of ratings from application contexts different from the one of the system’s intended one;
- only one system (Altered Vista) has been thoroughly evaluated through empirical studies that involved human users; and
- only one system (Evolving e-learning system) has been evaluated through a simulation that used a data set of ratings from the actual application context to which the system is intended to be deployed.
Table 3. An overview of selected recommender systems for learning resources, and related evaluation studies.

<table>
<thead>
<tr>
<th>System</th>
<th>Status</th>
<th>Evaluation</th>
<th>Focus</th>
<th>Type</th>
<th>Actors</th>
<th>Input</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altered Vista (Recker &amp; Wiley 2000; Recker &amp; Walker 2003; Recker et al. 2003)</td>
<td>Full system</td>
<td>Interface, algorithm, system usage</td>
<td>Pilot experiment, empirical study</td>
<td>Human users</td>
<td>User feedback</td>
<td>Overall satisfaction, usability, algorithm performance, social network formation</td>
<td></td>
</tr>
<tr>
<td>RACOFI (Anderson et al. 2003; Lemire et al. 2005)</td>
<td>Prototype</td>
<td>Algorithm</td>
<td>Simulation</td>
<td>System designers</td>
<td>Public data sets</td>
<td>Accuracy, complexity</td>
<td></td>
</tr>
<tr>
<td>QSAI (Rafaeli et al. 2004, 2005)</td>
<td>Full system</td>
<td>Algorithm</td>
<td>Simulation</td>
<td>System designers</td>
<td>Public data set</td>
<td>Accuracy</td>
<td></td>
</tr>
<tr>
<td>CYCLADES (Avancini &amp; Straccia 2005)</td>
<td>Full system</td>
<td>Algorithm</td>
<td>Simulation</td>
<td>System designers</td>
<td>Public data set</td>
<td>Overall satisfaction</td>
<td></td>
</tr>
<tr>
<td>Learning object sequencing (Shen &amp; Shen 2004)</td>
<td>Prototype</td>
<td>System usage</td>
<td>Pilot experiment</td>
<td>Human users</td>
<td>User feedback</td>
<td>Overall satisfaction</td>
<td></td>
</tr>
<tr>
<td>Evolving e-learning system (Tang &amp; McCalla 2003, 2004a,b,c, 2005)</td>
<td>Full system</td>
<td>Algorithm, system usage</td>
<td>Simulation, pilot experiment</td>
<td>Simulated users, human users</td>
<td>Actual data set</td>
<td>Accuracy, overall satisfaction</td>
<td></td>
</tr>
<tr>
<td>Combined Personalized Recommender System (Drachsler et al. 2007)</td>
<td>Prototype</td>
<td>Algorithm</td>
<td>Simulation</td>
<td>System designers</td>
<td>Actual data set</td>
<td>Accuracy</td>
<td></td>
</tr>
<tr>
<td>Multi-attribute Recommendation Service (Manouselis et al. 2007)</td>
<td>Prototype</td>
<td>Algorithm</td>
<td>Simulation</td>
<td>System designers</td>
<td>Actual data set</td>
<td>Accuracy, coverage, time</td>
<td></td>
</tr>
<tr>
<td>Learning Object Recommendation Model (Tsai et al. 2006)</td>
<td>Design</td>
<td></td>
<td></td>
<td>System designers</td>
<td>Actual data set</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>RecoSearch (Fiaidhi 2004)</td>
<td>Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>
Similar conclusions have also been found in a recent survey of such systems (Manouselis et al. 2010). A general observation that can be noted from Table 1 is that many times, the evaluation of proposed systems does not consider the actual requirements of the particular communities that they aim to serve. As a consequence, this may lead to incomplete or misleading conclusions about their performance. For instance, a recommendation algorithm that seems to be performing excellently on public data sets with ratings on movies or jokes (such as the popular MovieLens and Jester data sets) may demonstrate a different performance in the context of recommending the learning resources of an online repository such as MERLOT (Multimedia Educational Resource for Learning and Online Teaching, http://www.merlot.org).

Additionally, in many cases, a thorough experimental investigation of the recommendation algorithms does not take place. This is a common evaluation practice in systems examined for other domains (e.g. Breese et al. 1998; Herlocker et al. 2002; Deshpande & Karypis 2004; Papagelis & Plexousakis 2005). Previous evaluation studies have clearly indicated that careful testing and parameterization has to be carried out before a recommender system is finally deployed in a real setting. In this study, we perform such an experimental investigation of a candidate recommendation algorithm, by using real data from the intended application context.

Case study: EUN’s online community of teachers

EUN has carried out work on learning resources interoperability since 1999. This endeavour has resulted in the concept of the Learning Resources Exchange (LRE) (Van Assche & Massart 2004; Massart 2009), 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 to easily locate, use and reuse both open content as well as content from commercial suppliers.

EUN and supported online communities

The basic entrance to EUN’s resources is the LRE Home Page (Fig 1). A large percentage of its users often participate in various pilot actions that EUN is involved in, and they come from schools that are carefully selected by the local authorities in the participating countries. Thus, there is a better picture of their information and communication technology (ICT) maturity and their pedagogical practices, which are further enhanced by their participation in these initiatives. The users from these schools usually have good knowledge of computers, sufficient access to the Internet as well as sufficient pedagogical skills for the educational of digital learning resources. Very often, there already exist communities of practices that have been built around main interest areas despite geographical boundaries and which are further supported by such kind of initiatives. These are well-defined user samples, which makes it easier for EUN to execute pilot studies with their participation. On the other hand, users who come from open school collaboration projects use the portal usually without any formal engagements. They are usually driven to the portal though their own interest and willingness to get involved in European-wide collaborations in areas of interest, e.g. in science education. In some cases, there is some local promotion and support available from the collaboration projects. These users’ skills usually vary more than in the first case. However, they are in many cases the more innovative and leading-edge members of their own school communities.

Pilot study

The online community that has been the focus of the pilot study of this paper is the one of the CELEBRATE project. This project addressed all parts of the educational content value chain and involved 23 participants including Ministries of Education, universities, educational publishers, content developers, vendors and technology suppliers from 11 European countries. The aims of CELEBRATE were many-fold. On one hand, there was the interest to find out whether a learning resources’ brokerage system such as the one developed in the project allows ministries, publishers and users from individual schools to more easily exchange or sell 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 activities.
In the context of the project, an online K-12 teacher community of 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 1400 learning objects (i.e. digital resources that can be reused to support teaching and learning) and approximately 2400 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 the 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 of the outcomes of the pilot study has been the collection of a large number of multi-attribute evaluations of the particular resources that the teachers have

Fig 1 The learning resource exchange portal (http://lreforschools.eun.org).
accessed, viewed and in most cases, used as well. The questionnaire that has been used for this purpose included three sets of questions. The first dealt with assessing the potential classroom integration of the examined learning resource. The second was concerned with the ease of use of the learning resources and the third had been some demographics about the intended use of the learning resource. For the first two sets of questions, the teachers were asked to evaluate the learning resources upon several attributes using the 1–5 evaluation scale (strongly disagree, disagree, neither disagree nor agree, agree, strongly agree) (Fig 2). These attributes were ease to integrate resource in classroom work, relevance to teaching topics, facilitate students’ learning, ease of use, not much change/adaptation for classroom use and use with variety of teaching approaches.

The motivation behind their selection was identifying criteria that are particularly meaningful for the teachers (i.e. the users of the specific portal) when searching for and selecting a resource for use in the classroom. For the first two sets of questions, the teachers were asked to evaluate the learning resources upon several attributes using the 1–5 evaluation scale (strongly disagree, disagree, neither disagree nor agree, agree, strongly agree) (Fig 2). These attributes were ease to integrate resource in classroom work, relevance to teaching topics, facilitate students’ learning, ease of use, not much change/adaptation for classroom use and use with variety of teaching approaches.

The pilot study led to the collection of an evaluations’ data set that has been judged as particularly useful for testing a collaborative filtering service before it is deployed to support this community. Data sets with users’ feedback, such as the widely known MovieLens and EachMovie data sets (Herlocker et al. 2004; Maritza et al. 2004), 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 (Tso & Schmidt-Thieme 2006; Manouselis & Costopoulou 2007a). Thus, we have decided to use this data set in order to experimentally investigate various design options for a proposed collaborative filtering algorithm that takes multi-attribute evaluations as input. The data set included 2554 evaluations related to 899 learning resources in LRE’s repositories, which have been provided by 228 teachers. The average number of ratings per user was 11.2.

Experimental investigation of collaborative filtering for the LRE portal

In this section, we introduce the collaborative filtering algorithm considered for implementation in the LRE portal, and we present how CollaFiS, a simulation environment for collaborative filtering algorithms (Manouselis & Costopoulou 2006), has been used to investigate several design options for them. To assume conditions similar to the ones expected during the actual operation of the system, the data set collected by the pilot study described in the previous section has been used. The results of this experiment provide a useful insight into how the examined algorithm will perform when deployed to support the particular online community.

Examined algorithm

The goal of collaborative filtering service will be to provide to some member of the online community (which corresponds to the active user \( a \in C \)) with an estimation of how he or she would evaluate a particular target item \( s \) that he or she has not previously seen or with a recommended ranking of items that he or she has not previously seen, which he or she would appreciate higher than the others. To calculate this prediction, a neighbourhood-based collaborative filtering algorithm was adopted. Neighbourhood-based algorithms are the most prevalent approaches for collaborative filtering (Herlocker et al. 2002; Zeng et al. 2004). They create a neighbourhood \( D \subseteq C \) 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 neighbourhood have evaluated \( s \).
The studied algorithm was a multi-attribute extension of related algorithms, which is based on Multi-Attribute Utility Theory (Keeney 1992). It considers each attribute in separate, first trying to predict how the active user would evaluate it upon each attribute and then synthesizing these attribute-based predictions into a total utility value. A variety of design options can be considered for the studied algorithm, leading to several versions and parameterizations. A detailed explanation of these options can be found in Manouselis and Costopoulou (2007a). In this paper, we illustrate how taking advantage of the collected data set with teacher evaluations, we experimented with more than 360 algorithm variations in order to select the most appropriate one for the EUN portal. A detailed analysis of the mechanics behind this experiment can be found in Manouselis et al. (2007).

**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 portal. A number of design options have been examined (Manouselis et al. 2007). The CollaFiS simulation environment (Manouselis & Costopoulou 2006) 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 one testing component (using an 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 (2043 evaluations).

For our experimental testing, three particular performance evaluation metrics have been used:

1. **Accuracy**: predictive accuracy of the algorithms has been measured through the mean-absolute error (MAE) of the predicted utility against the actual utility of an item.

2. **Coverage**: coverage of the algorithms has been measured as the percentage of items for which an algorithm could produce a prediction.

3. **Time**: prediction speed has been measured 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 personal computer with a Pentium 4 (2.5 GhZ, 256 MB RAM; Intel Corporation, Santa Clara, CA) running Microsoft Windows XP (Microsoft Corporation, Redmond, WA), Apache server 1.3.33, PHP 5.0.3, and MySQL Server 4.1.

**Results**

As the detailed comparison results of all algorithm variations has indicated (Manouselis et al. 2007), several algorithm variations could be used in the EUN portal with satisfactory performance. In order 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 examined their execution time. Table 4 demonstrates how 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 a particular algorithm variation that offers a combination of 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 s) for the studied data set.

**Discussion**

Using data gathered from user-generated ratings is a common means in recommending movies, music, books and commercial products. However, for a learning com-
Collaborative recommendation

Table 4. Top 10 algorithm variations according to MAE and execution time (with coverage >65%).

<table>
<thead>
<tr>
<th>Rank</th>
<th>Version</th>
<th>Neighborhood method</th>
<th>Normalization method</th>
<th>MAE</th>
<th>Coverage (%)</th>
<th>Execution time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>Cosine</td>
<td>MNN = 4</td>
<td>Deviation from mean</td>
<td>0.57028</td>
<td>69.08</td>
<td>27</td>
</tr>
<tr>
<td>2nd</td>
<td>Cosine</td>
<td>MNN = 1</td>
<td>Deviation from mean</td>
<td>0.59111</td>
<td>69.08</td>
<td>23</td>
</tr>
<tr>
<td>3rd</td>
<td>Cosine</td>
<td>CWT = 0.85</td>
<td>Simple mean</td>
<td>0.65388</td>
<td>63.80</td>
<td>17</td>
</tr>
<tr>
<td>4th</td>
<td>Cosine</td>
<td>CWT = 0.85</td>
<td>Weighted mean</td>
<td>0.65390</td>
<td>63.80</td>
<td>17</td>
</tr>
<tr>
<td>5th</td>
<td>Euclidian</td>
<td>MNN = 3</td>
<td>Simple mean</td>
<td>0.67257</td>
<td>69.08</td>
<td>19</td>
</tr>
<tr>
<td>6th</td>
<td>Pearson</td>
<td>MNN = 6</td>
<td>Simple mean</td>
<td>0.67553</td>
<td>69.08</td>
<td>22</td>
</tr>
<tr>
<td>7th</td>
<td>Cosine</td>
<td>CWT = 0.55</td>
<td>Weighted mean</td>
<td>0.67650</td>
<td>69.08</td>
<td>17</td>
</tr>
<tr>
<td>8th</td>
<td>Euclidian</td>
<td>MNN = 9</td>
<td>Simple mean</td>
<td>0.67682</td>
<td>69.08</td>
<td>19</td>
</tr>
<tr>
<td>9th</td>
<td>Pearson</td>
<td>MNN = 8</td>
<td>Simple mean</td>
<td>0.67685</td>
<td>69.08</td>
<td>21</td>
</tr>
<tr>
<td>10th</td>
<td>Euclidian</td>
<td>MNN = 14</td>
<td>Simple mean</td>
<td>0.67718</td>
<td>69.08</td>
<td>18</td>
</tr>
</tbody>
</table>

An explanation of what each design option means (i.e., Version, Neighbourhood Method, Normalization Method) is provided in Manouselis et al. (2007).

Collaborative recommendation

Community, it is not clear whether the users would find recommendations, based on similarities in rating patterns, useful. As much as emphasis is given to parameterization of the data and respective algorithms, the question of users in the centre has to be asked. It has been argued that focusing only on ‘technical’ evaluation is not enough for recommender systems: often, it is carried out without considering the actual needs and the opinion of the end-user (Swearingen & Sinha 2001; Herlocker et al. 2004; McNee et al. 2006). Thus, further evaluation strategies that will work complementary to technical evaluations need to be applied in order to bridge the gap between users, their information-seeking tasks and the deployed systems. For instance, human-centred evaluation approaches, such as the Human-Recommender Interaction (HRI) framework that has been proposed by McNee (2006) may be scrutinized. The HRI framework focuses on the suitability of a given algorithm to the information-seeking task that the user has at hand and could help us evaluate whether the proposed collaborative filtering system actually helps the community members achieve their information-seeking tasks.

Moreover, it will be important to ask what type of data would be relevant for a teachers’ community in order to generate recommendations. The data gathered from users’ evaluations on resources are intensive to produce since not all the users contribute explicit ratings. In addition, a discussion about the appropriateness (e.g. completeness and exhaustiveness) of the evaluated attributes and the engaged evaluation scale could raise several objections to the studied approach. Further exploration on the type and amount of data that can be collected (explicitly or implicitly) by the members of the community is required. Adomavicius et al. (2005) discussed about the way input data are collected for a recommender system whether it is performed implicitly (e.g. previous actions) or explicitly (e.g. ratings). They argue for a more implicit data gathering in order to support the recommendation process in a more sustainable and non-intrusive manner. In the same line of thought, EUN is investigating towards the direction of tracking user actions by logging attention-related metadata (Najjar et al. 2006) and aims to conduct further studies on what type of data are most suitable for its communities’ purpose. A hypothetic argument could be that data gathered from previous resources bookmarked and used (e.g. domain area, intended audience, language) could yield important information when combined with rating data whenever available.

In addition, related to the issues of data sources for a recommender is the question of the information-seeking patterns across tasks and contexts that teachers and learners are involved in when they address the LRE portal and how could they be supported by social IR methods. Information seeking, similar to learning, is a fundamental and high-level cognitive process (Marchionini 1995) and in some cases, can be a social activity. Hargittai and Hinnant (2006) argue that an ‘important factor influencing users’ information-seeking behaviour concerns the availability of social support networks to help address users’ needs and interests. People’s information behaviour does not happen in isolation of others.’ Thus, more knowledge is needed on teachers’ information-seeking tasks, which can be contextually driven, i.e. there is the national or regional curriculum but on the other hand, can have many...
common points on the European level, such as topics, learning goals and activities that can be shared. Another important perspective is the exploration of existing and emerging social ties between the members of online communities and their exploitation for modeling the social networks. In this way, rating-based collaborative filtering of learning resources can be further supported by information stemming from the analysis of the network’s social relationships (Domingos 2005). These can be personal evaluation histories (Vuorikari et al. 2008) and information regarding social book marking and tagging.

To conclude the discussion on this simulation experiment, it is important to acknowledge the aforementioned points and understand that as opposed to movie or music recommender system, the item and problem base in the field of learning resources is very diverse and rather complex because of the variety in the potential educational uses of a learning resource. Technical-oriented experiments such as the one presented in this paper are particularly valuable but need to be further complemented with studies that will focus on the pedagogical/educational value of recommender systems, e.g. for collaborative learning activities (Recker et al. 2003) and for enhancing the social perspective of learning (Rafaeli et al. 2004, 2005). It can be even argued that educational uses of collaborative filtering should go beyond simple good/bad ratings as it usually happens in traditional recommender systems. Because learning needs and capacities change and evolve, it may be not safe to assume that past equivalence to assume future equivalence. So expressing preferences for learning is not similar to preferences expressed for movies or books, and there is a need to develop models and algorithms that go beyond neighbourhood-based approaches and can try to identify and predict trajectories of change. A discussion on such challenges took place in Manouselis et al. (2010). Nevertheless, the work already accomplished in the field of recommender systems for education indicates that even simple collaborative filtering approaches might provide some value to the users of portals with learning resources (such as the EUN one).

Conclusions

In this paper, we introduced the concept of collaborative filtering of learning resources and reviewed characteristic approaches. We also highlighted a typical problem when designing such systems: the fact that testing takes place using data from other environments/communities than the one that the system is being built to support. To this end, we have illustrated how we used collected data from the teachers of the EUN portal in order to experimentally investigate how different variations of a particular multi-attribute collaborative filtering algorithm perform. In this way, it became possible to choose a particular parameterization that seems as the most appropriate for the online community of this portal.

This simulation experiment may serve as a first step towards the understanding and appropriate specialization for a collaborative filtering service for the given user community. We aim to further complement it with experiments studying 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 (McNee 2006). The design variables and pedagogical conditions for a social recommender system that aims to enhance the use and reuse of learning resources constitute a complex problem space where many stakeholders and their needs need to be covered. Although technology is an enabler of the process, the user should stand in the centre with his or hers educational specific needs and information seeking tasks at hand.

Acknowledgements

Part of this work has been funded with support by the European Commission and more specifically, the projects ‘MELT: a Metadata Ecology for Learning and Teaching’ and ‘Organic.Edunet: A Multilingual Federation of Learning Repositories with Quality Content for the Awareness and Education of European Youth about Organic Agriculture and Agroecology’ of the eContentplus Programme.

References


Adomavicius G. & Tuzhilin A. (2005) 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**, 734–749.


Gomez-Albarran M. & Jimenez-Diaz G. (2009) Recommendation and students’ authoring in repositories of learning...


