

Recommendation of Collaborative Activities in E-learning Environments

Pierpaolo Di Bitonto, Maria Laterza, Teresa Roselli, and Veronica Rossano

Department of Computer Science, University of Bari,
Via Orabona, 4 – 70125 Bari – Italy
dibitonto@di.uniba.it,
{teresa.roselli, veronica.rossano}@uniba.it

Abstract. In distance education environments, collaborative activities such as wikis, forums and chats play an important role in the e-learning experience because they promote communication among students and so allow cooperative learning settings to be implemented. Nevertheless, it could be difficult for learners to pick out the most interesting and appropriate collaborative activities to meet their learning needs. Recommender systems integrated in e-learning platforms are usually used mainly to help learners choose teaching resources, but they can also be useful to suggest the collaborative activities that best fit their learning objectives from a pedagogical point of view. In this context, the paper presents a recommendation approach able to suggest collaborative activities such as forums, chats, wikis and blogs, that combines dynamic clustering and prediction calculus on the basis of the learners' profiles and needs.

Keywords: Recommender system, collaborative learning, dynamic clustering.

1 Introduction

Collaborative activities, such as chats, forums, blogs, wikis, file sharing, are becoming increasingly popular in the field of Education because they are helpful tools that support both teaching and learning activities. In particular, they allow cooperative learning settings to be implemented in distance education, that foster deeper reflection and mutual growth processes in students. Although these activities promote communication among students, which is one of the basic requirements for a successful e-learning experience, it could be difficult for the learner to pick out the most interesting and appropriate activities for her/his learning needs. Ideally, e-learning platforms should be able to help students to address their information overload, suggesting the best collaborative activity to tackle next according to her/his preferred learning style, cognitive and metacognitive abilities, interests, and learning needs.

Recommender systems have taken on a leading role in many applications (such as in e-commerce, e-government, e-learning, etc.) where users have too many choices, too little time, and in which the information explosion makes the problem even more

difficult. A recommender system integrated in an e-learning platform could help learners not only to choose the best teaching resource, as usually happens, but also to suggest the best collaborative activity for her/his learning needs. Up to now great numbers of recommender systems in e-learning have been developed to suggest courses, learning materials [1], as well as relevant topics in a forum [2] but only few researches have combined clustering with recommendation techniques for suggesting collaborative activities to stimulate and support cooperative and collaborative learning [3].

The paper presents a recommendation approach integrated in an open source e-learning platform, designed to suggest collaborative activities such as forums, chats, wikis and blogs. The approach combines dynamic clustering of learners on the basis of their learning needs and prediction calculus of the most suitable new collaborative activities (in the sense that the specific user has not yet been involved in them) for the current learner on the basis of similar users' interests.

The paper is structured as follows: section 2 presents examples of recommendation systems in e-learning settings; section 3 provides a brief description of the proposed recommendation approach; section 4 illustrates an example of interaction on the Moodle platform; section 5 reports a preliminary evaluation of the recommendation approach's effectiveness and, finally, section 5 outlines some conclusions and future works.

2 Related Works

Nowadays, one of the most interesting challenges in the e-learning research field is to address the problem of information overload. Most of the existing recommender systems merely suggest educational resources such as: learning objects (LOs), simulations, demos, exercises and so forth. Altered Vista [13], for example, suggests web-based teaching resources to both students and teachers on the basis of a collaborative user-based method [14] and the multidimensional users' opinions. In particular, users can judge the usefulness of the resource, the accuracy of the information, the educational relevance for a specific learning objective and so on. Also the RACOFI system (Rule-Appling Collaborative Filtering system) [15], which suggests audio LOs via a collaborative technique, uses multidimensional ratings to evaluate LOs and to predict evaluations of the new resources that best fit the user queries.

In the Web 2.0 era, an e-learning experience cannot be limited to using the Net to convey learning materials, because it is also important to promote communication among students using collaborative activities such as chats, forums, blogs and wikis, which are one of the essential factors for a successful e-learning experience. Although they are becoming increasingly popular in Education, up to now only few researches have aimed to suggest collaborative learning resources. Among the examples of research in this direction there is the Personal Recommender System (PRS) [16], which suggests the most suitable learning activities to help the specific learner improve her/his learning path. In particular, the PRS employs a switching

hybridization strategy in order to identify which is the most suitable prediction technique, between knowledge-based [17] and collaborative with demographic features. In other words, if only the student's interests are available, the system uses a knowledge-based technique; otherwise, if further information is available (motivation, time spent studying and so on) it uses the collaborative technique.

However, the use of merely recommender system methods, either pure or hybrid, is not sufficient to make effective suggestions. Usually, it is fruitful to combine the recommendation approach and web mining techniques. An example is Zaiane's research [18] that recommends, using the agents approach and web mining techniques, on-line learning activities or shortcuts in web-based learning on the basis of the user's access. The proposed e-learning recommender system, in fact, by using the association of rule mining, analyzes text messages sent during a specific discussion, identifies the context or theme of the discussion and offers the participants suggestions about exercises, previous users' posted messages on a conferencing system, on-line simulations, or simply a resource in the web. Another example is the recommender system described in Tang and McCalla [3], that suggests learning activities using clustering techniques. In particular, the proposed recommender approach firstly groups learners according to their learning interests and background knowledge using a clustering technique, then makes predictions based on the users' explicit ratings by a collaborative filtering technique.

Tang and McCalla stated that applying clustering before the prediction is made allows collaborative filtering to be guided towards obtaining more accurate recommendations. In accordance with this statement, our aim is to suggest collaborative activities to stimulate and support cooperative learning using clustering and collaborative recommendation techniques, but unlike the Tang and McCalla solution, the proposed approach combines dynamic clustering and collaborative filtering in order to predict the most suitable new collaborative activities (in the sense that the user has not yet been involved in them) for the current learner on the basis of similar users' interests.

3 The Recommendation Approach

Usually, a Recommender System is used to identify sets of items that are likely to be of interest to a certain user, exploiting a variety of information sources related to the user, the content items and perhaps also the recommendation context. The recommendation process starts with the specification of the initial set of data that is either explicitly provided by the user or is implicitly inferred by the system [9].

Over the past two decades many different pure or hybrid [10] algorithmic approaches have been proposed both to select and to rank the item set to be suggested. These algorithms can also be classified as *heuristic-* or *memory-based* and *model-based* according to the formulation of the function R [11].

The first two require all ratings, items, and user data to be stored in the memory and estimate the utility of each item for the current user using all the data collected, based on a certain heuristic assumption. For example, the assumption that two users who show similar preferences for already-seen items will have similar preferences for

unseen items as well. The latter approach learns a predictive model offline (using statistical or machine learning methods) and uses this model to compute the function R for unseen items [12].

Collaborative recommendation algorithms are probably the most familiar and most widely implemented in many domains [10]. Beyond the data gathering and storing method, these algorithms aggregate ratings or recommendations of objects, recognize common features among users on the basis of their ratings, and generate new recommendations based on inter-user comparisons [10].

In particular, collaborative filtering is one of the most popular recommendation methods used in e-learning environments, because it estimates the utility of a teaching resource for a user on the basis of the ratings attributed to that particular resource by the community. In other words, a collaborative algorithm uses ratings on items already seen both by the active user and other system users to predict ratings for unseen items.

Despite the fact that the collaborative method has always been considered one of the best recommendation methods to use in e-learning, it has been proven that when it is combined with AI approaches the predictions can be more accurate and more effective recommendations are supplied to users. In particular, clustering techniques may be advantageous for group users (students or teachers) of an e-learning platform as a means of offering personalized learning contents, paths and activities. But study of the state of the art demonstrates that the use of clustering techniques is not yet widespread. This is probably due to the fact that the most popular clustering algorithms require the number of clusters in which the users will be divided to be known a priori, which is not the most efficient approach for use in an e-learning community where users and preferences evolve continuously.

The proposal presented here addresses this problem, achieving dynamic clustering of learners by combining the Silhouette index and K-means algorithm.

3.1 The Dynamic Clustering of Learners

The main aim of the proposed algorithmic approach is to suggest collaborative activities such as forums, chats, wikis and blogs in an e-learning environment. In particular, the defined approach combines *dynamic clustering*, that clusters the learner on the basis of her/his learning needs, with the prediction of the most suitable new collaborative activities to be proposed on the basis of the interests of similar users to the target one.

In general, the main goal of clustering is to group objects into clusters so that the objects in the same cluster are more similar to each other than to those in other clusters. One of the most commonly used clustering algorithms in recommender systems is the K-means [4], which classifies a given data set in a number of clusters specified in advance. In the e-learning context, where the learning dimensions (preferences, goal, background knowledge, etc.) have a high level of dynamism, the need to specify a priori the number of clusters is an important drawback of this clustering method. For this reason, the proposed solution uses dynamic clustering [6], which classifies learners on the basis of similar learning needs and interests, without requiring an initial indication of the number of clusters. In particular, the proposed approach uses the *Silhouette index* [5] to estimate the optimal number of clusters in

which to group the data set and the *K-means algorithm* [4] to cluster the data set into the optimal, previously defined partition.

Let us consider a data set with n samples that has to be divided into k clusters, whereby the Silhouette index will be:

$$Sil(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (1)$$

where $a(i)$ is the average dissimilarity of sample i to all other samples within the same cluster computed using a distance measure; $b(i)$ is the lowest average dissimilarity of i to any other cluster. In other words, the lower the values of both $a(i)$ and $b(i)$ the better the match. Function $Sil(i)$ ranges from -1 to 1; when it is close to 1 it means that the sample is appropriately clustered; if it is close to 0 it means that the sample is on the border of two natural clusters; otherwise if it is close to -1 it is misclassified.

In the e-learning context the sample is composed of learners, and each learner will be represented by a vector describing her/his learning needs inferred in several ways according to the available tracking data in the e-learning platform, such as the assessment test results, time spent by the learner on a teaching resource, and so on.

Using the formula (1), in an iterative procedure, it is possible to calculate several values of Sil corresponding to different clustering solutions. The optimal number of clusters corresponds to the largest value of the *Silhouette index*. The number of clusters calculated is the input of the K-means algorithm, which aims to classify the sample of learners in the defined number of clusters [19].

When the clustering process is completed the data set is partitioned in several clusters of users, so that the learners in the same cluster have similar learning needs. In each cluster there is a center, named *centroid*, that is used during the prediction calculus in order to identify the neighbourhood of the target learner. The neighbourhood is defined by computing the similarity of two clusters, i.e. the similarity of their centroids. Those clusters that have the highest Pearson correlation coefficient from the centroid of the target learner cluster is the learner's neighbourhood. Thus, the prediction calculus step computes a score for all the collaborative activities in which the target learner has not yet been involved, on the basis of the interests showed by the users belonging to the learner's neighbourhood. In other words, the learner's favourite collaborative activities are inferred through the rate of interventions of pairs that have similar learning needs and interests in a collaborative tool (chat, forum, wiki or blog).

The proposed recommender approach can be easily implemented and integrated in any e-learning platform, in order to suggest the best collaborative learning activity for the learner. In particular, we have implemented a plug-in for one of the most commonly used open source e-learning platforms: Moodle

4 Example of Interaction in Moodle

In order to validate the proposed recommendation system a plug-in in Moodle has been implemented, which allows forums, chats, and wikis to be suggested. The plug-in architecture consists of three components (Fig.1):

- The *Data Gathering Component* (DGC) analyses and extracts the tracking data from the Moodle DB. As described above, each learner is described by using a vector in which learning needs and interests are coded. The learning needs are inferred using data about her/his participation in courses and test results. The learner's interests are inferred using her/his participation in forums, chats and wikis. In particular, as regards interest in a forum, it is calculated as the percentage of the user's posts, in the forums of a specific course, out of the total number of posts since her/his first intervention. For chat the measure of interest is the user's participation in different chats about a course, in other words the percentage is calculated on the number of chats in which the learner has taken part, out of the total number of chats in the course. The same process has been used for wikis, whereby the percentage is calculated as the number of actions in the shared documents out of the total number of actions;
- The *Clustering Component* (CC) clusters the users using the data collected by the DGC. Firstly, using the Silhouette index it defines the optimal number of clusters, then by using the K-means algorithm it classifies learners in the defined number of clusters. The output of this component is the definition of classes of users with similar learning needs;
- The *Prediction Component* (PC) computes a score for all the collaborative activities in which the user has not yet been involved. Firstly, the target learner neighbourhood is identified by computing the Pearson correlation coefficient [7] among this and all the other cluster centroids; then, using the data about the interest in forums, chats and wikis of neighbours, the prediction on the unseen collaborative activities is calculated.

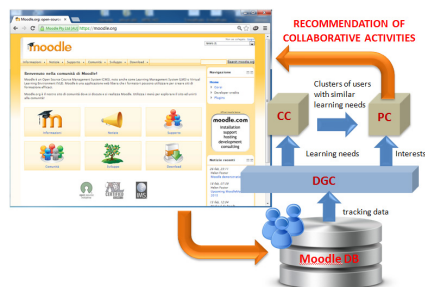


Fig. 1. Recommender system components

5 Results and Discussions

There are already a large number of studies offering empirical evidence that users prefer a recommender system that provides more accurate predictions. In the e-learning domain, we assumed that a recommender is successful if it is able to provide accurate predictions and to promote a deeper level of learning. Thus, for the defined approach it is necessary to measure both the prediction accuracy. In our pilot study we measured the prediction accuracy only for the forums, because they were the

collaborative activity used in the specific experimental context. In particular, the prediction accuracy has been evaluated using the popular precision and recall metrics [8], where the precision is the fraction of retrieved items that are relevant to the search, while recall is the fraction of relevant items that are retrieved by query. A forum was assumed to be relevant for the target user if after the recommendation the learner decided to participate in it.

In order to have a meaningful sample to measure the quality of the suggestions supplied, students and forums focusing on different courses at the Multimedial Advanced Educational Technology Laboratory of the Science Faculties at the University of Bari were enrolled. This Laboratory offers e-learning services to different degree courses of the Science Faculty. In particular, the recommendation system has been applied in two degree courses on Informatics and Digital Communication (located in Bari and Taranto). The sample considered consists of 234 students, 10 courses and 26 forums.

The recall points were fixed and the precision was calculated for each of them. Because precision values were lacking for some standard recall points (0%, 10%, 20%...,100%), interpolated precision values were computed. The average precision values were computed considering 50 recommendation requests corresponding to standard recall levels.

Finally, a precision-recall curve was drawn. It emphasizes the proportion of preferred items that were actually recommended (Fig.2).

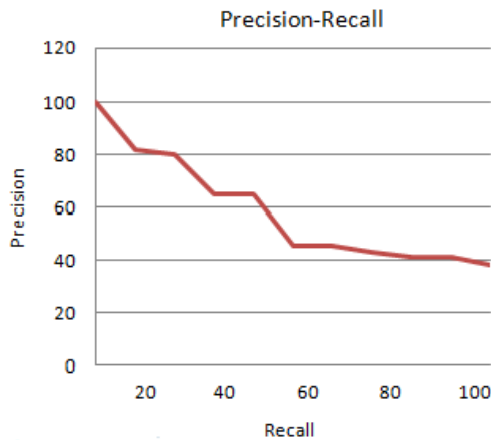


Fig. 2. Precision-recall curve

Results showed a high prediction accuracy mainly for low recall values (up to 30%). Average values of precision ranged between 20% and 50% of recall. Finally, the precision value is lower when recall is about 60% and remains constant up to 100%. Moreover, the curve is very distant from the origin of the axes. To summarize, the recommender system shows a good performance and supplies accurate predictions.

6 Conclusions and Future Works

Collaborative activities are very helpful tools in e-learning environments because they promote cooperative learning and deeper reflection. However, due to the great number of activities in which the learner could be involved in an e-learning platform, it may be difficult for her/him to choose the most appropriate activities for her/his learning needs and interests.

This problem can be solved by using recommendation systems integrated in e-learning platforms that suggest the best collaborative activities according to the user's preferred learning style, cognitive and metacognitive abilities, interests, and learning needs.

The paper illustrates the added value of combining dynamic clustering with collaborative filtering. The proposed solution, in fact, is a two step approach that combines the *dynamic clustering* of the learner on the basis of her/his learning needs and interests, and the *prediction calculus* of the most suitable collaborative activity (in which the user has not yet been involved) for the current user according to the favourite collaborative activities of users with similar learning needs and interests.

The defined recommender approach can be easily implemented and integrated in any e-learning platform, in order to suggest the collaborative learning activities best suited to each learner. In order to evaluate our approach, we have implemented it in a plug-in for one of the most commonly used open source e-learning platforms: Moodle. The approach was evaluated using data gathered from interactions of students on different courses in the Multimedial Advanced Educational Technology Laboratory of the Science Faculties at the University of Bari. Based on the results obtained, it is clear that the accuracy of the recommendation is very high. An evaluation aimed at comparing the recommendation accuracy of the proposed approach with traditional recommendation algorithms is currently being planned.

References

1. Wang, P.Y.: The Analysis and Design of Educational Recommender Systems. In: Carlsen, R., et al. (eds.) Proceedings of Society for Information Technology & Teacher Education International Conference 2007, pp. 2134–2140. AACE, Chesapeake (2007)
2. Castro-Herrera, C., Cleland-Huang, J., Mobasher, B.: A recommender system for dynamically evolving online forums. In: Proceedings of the Third ACM Conference on Recommender systems (RecSys 2009), pp. 213–216. ACM, New York (2009)
3. Tang, T., McCalla, G.: Smart recommendation for an evolving e-learning system. *International Journal on E-Learning* 4(1), 105–129 (2005)
4. Steinbach, M., Karypis, G., Kumar, V.: A comparison of document clustering techniques. In: *KDD Workshop on Text Mining*, vol. 400, pp. 525–526 (2000)
5. Jaafar, A., Sareni, B., Roboam, X.: Clustering analysis of railway driving missions with niching. *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering* 31(3), 920–931 (2012)

6. Chiang, M.M.-T., Mirkin, B.: Experiments for the Number of Clusters in K-Means. In: Neves, J., Santos, M.F., Machado, J.M. (eds.) EPIA 2007. LNCS (LNAI), vol. 4874, pp. 395–405. Springer, Heidelberg (2007)
7. Ben Schafer, J., Frankowski, D., Herlocker, J., Sen, S.: Collaborative Filtering Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007)
8. Shani, G., Gunawardana, A.: Evaluating recommendation systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 257–298. Springer (2011)
9. Adomavicius, G., Kwon, Y.: New Recommendation Techniques for Multi-Criteria Rating Systems. *IEEE Intelligent Systems* 22, 48–55 (2007)
10. Burke, R.: Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction* 12(4), 331–370 (2002) ISSN: 0924-1868, doi:10.1023/A:1021240730564
11. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.): *Recommender Systems Handbook*. Springer Science + Business Media, New York (2011) ISBN/ISSN: 9780387858197 0387858199, doi: 10.1007/978-0-387-85820-3_2, LLC 201
12. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, pp. 43–52 (1998)
13. Walker, A., Recker, M., Lawless, K., Wiley, D.: Collaborative information filtering: A review and an educational application. *International Journal of Artificial Intelligence and Education* 14, 3–28 (2004)
14. Ben Schafer, J., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007)
15. Anderson, M., Ball, M., Boley, H., Greene, S., Howse, N., Lemire, D., McGrath, S.: RACOFI: a Rule-Appling Collaborative Filtering System. *IEEE/WIC*, Halifax (2003)
16. Drachler, H., Hummel, H.G.K., Van den Berg, B., Eshuis, J., Berlanga, A.J., Nadolski, R.J., Waterink, W., Boers, N., Koper, R.: Recommendation strategies for e-learning: preliminary effects of a personal recommender system for lifelong learners, Maastricht (2007)
17. Burke, R.: Knowledge-based Recommender Systems. In: Kent, A. (ed.) *Encyclopedia of Library and Information Systems*, Marcel Dekker, New York (2000)
18. Zaïane, O.: Building a recommender agent for e-learning systems. In: ICCE, pp. 55–59 (2002)
19. Wang, K., Zhang, J., Li, D., Zhang, X., Guo, T.: Adaptive Affinity Propagation Clustering. *Acta Automatica Sinica* 33(12)