e-Learning Experience using Recommender Systems

Jesus Bobadilla, Antonio Hernando, Angel Arroyo
Universidad Politécnica de Madrid, Crta. De Valencia, Km. 7
28031 Madrid, Spain
+34 913365133
jesus.bobadilla@upm.es, ahernando@eui.upm.es, aarroyo@eui.upm.es

ABSTRACT
This paper presents the results obtained using a real e-learning recommender system where the collaborative filtering core has been adapted with the aim of weighting the importance of the recommendations in accordance with the users’ knowledge. In this way, ratings from users with better knowledge of the given subject will have greater importance over ratings from users with less knowledge.

In the same way, we validate the results obtained and we adjust, with just one parameter, the weight that should be awarded, in each specific e-learning recommender system, to the ratings of the users with the best reputation.

The results obtained show a notable improvement regarding traditional collaborative filtering methods and suggest balanced weightings between the importance assigned to users with more or less knowledge.

Categories and Subject Descriptors
K.3.2 [Computer and Information Science Education]

General Terms
Experimentation.

Keywords
Recommender Systems; e-learning; collaborative filtering.

1. INTRODUCTION
The social web is leading to positive changes in the way we approach many different aspects of our everyday lives. Nowadays, we often take into consideration the opinions, experiences, tastes and information provided by other people that we do not even know, but whom we turn to more and more when doing our shopping, planning our holidays, choosing the films we watch, selecting the news we read, etc. Of course, in general, we do not base our opinions on the information provided by just one user or even a few users, but rather, we immerse ourselves in a series of searches via one or more web 2.0 websites in order to obtain information that we did not know and that it would be hard to find on non-social websites (e.g. the story of the entertaining late-night chats with the owner of a house who rents out rooms in a national park in New Zealand). We just need to find various congruent comments in order to make a decision on whether to stay at that place or not and, in general, we will make the right decision and will add our experience and comments to reinforce the choice of that destination by future travelers.

Therefore, the social web helps us in our everyday lives as it enables us to access all kinds of information that is not necessarily institutional or structured, provided by any other person. It represents progress made in the existing possibilities regarding traditional websites; however, throughout their rapid expansion, social webs have come across the same problem as their non-social predecessors: sifting out suitable information from a huge and growing number of opinions, posts, photos, videos, texts, etc.

The tools most commonly used to search for the information that best adapts to our expectations are still the traditional search engines (Google, Yahoo,…); we can also see a social tendency to gather in social mega-webs (the most famous of which is Facebook) where we are much more likely to find what we are looking for. With addition of powerful technology compatible with the above-mentioned tools we have recommender systems (RS); At present, recommender systems (RS) comprise an important collaboration tool among Web 2.0 users [21,25,19], providing services that enable accurate recommendations to be made to internet users who have been previously defined according to their preferences in a certain field.

Nowadays, the most broadly-studied RS focus on recommending films [23,2,24] based on votes previously made by users, who have therefore defined their preferences on a limited number of these films. Nevertheless, RS have been introduced to a greater extent in the collaborative systems environment and they cover a large number of applications [4,12] in which it is useful to receive recommendations based on the preferences of a group of users with similar tastes or needs to those of each individual that makes use of these systems; in the commercial environment, the most widely demanded RS applications are those that refer to e-commerce [20,8,31].

Traditionally, papers published in the area of RS research have been based on studies of general validity, which can be equally applied to any type of RS and are normally based on results obtained using existing public databases (which usually contain film ratings). However, it is now possible to find studies focusing on different specific types of RS, such as e-learning RS [26,11,10,17,14,3].

The basic principle of RS is the expectation that the group of users similar to one given user, (i.e. those that have rated an important number of elements in a similar way to the user) can be used to adequately predict that individual’s ratings on products the user has no knowledge of. This way, a trip to Senegal could be recommended to an individual who has rated different destinations in the Caribbean very highly, based on the positive ratings about the
holiday destination of “Senegal” of an important number of individuals who also rated destinations in the Caribbean very highly. This suggestion (recommendation) will provide the user of the service with inspiring information from the collective knowledge of all other users of the service.

In short, the internal operating core of RS is based on carrying out collaborative filtering (CF) [1,16] starting from the ratings expressed by a group of users about a group of items, the aim is to select users who have the most similar ratings or tastes to those of the individual who is using the system at any one time. In general, the objective is to suggest a series of elements to the individual, on which this individual have not shown a preference but which have been very highly rated by an important proportion of the group of users with similar preferences to the individual.

The quality of the results offered by a RS greatly depends on the quality of the results provided by its CF [1] phase; i.e. it is essential to be capable of adequately selecting the group of users most similar to a given individual. Two main different approaches are available to tackle this important task: memory-based methods and model based methods.

Memory-based methods [7,22,30], use metrics [14,30] that are directly applied to the data matrix that contains the ratings made by the set of users of the system on the set of items available. The current RS for commercial use employ memory-based methods due to their robustness, predictability and efficiency.

Model based methods use data to create a model: bayesian classifier [9], neural network [18], fuzzy system [32], etc. and from this model they predict the clusters of similar users. Model based methods are generally found in the research stage, in the implementation of recommender systems for non-commercial use.

Although a great variety of metrics capable of implementing the CF have been published, most of the systems only use two ways of measuring the distance between users: cosine and Pearson correlation. This is due to the fact that these metrics have obtained the best results in various research projects and that their behavior in RS is satisfactory.

Regardless of the method used in the CF stage, the technical objective generally pursued is to minimize the prediction errors, by making the accuracy [15,28,13,27] of the RS as high as possible; nevertheless, there are other objectives that need to be taken into account: avoid overspecialization phenomena, find good items, credibility of recommendations, precision and recall measures, etc.

Memory-based methods work on a table of U users who have rated I items. The prediction of a non-rated item i for a user u is computed as an aggregate of the ratings of the K most similar users (k-neighborhoods) for the same item i, where \( \bar{K} \) denotes the set of k-neighborhoods.

### 2. COLLABORATIVE FILTERING PROPOSED FOR E-LEARNING

We propose a reflection on part of the existing differences, from the learning perspective, between a well-known non-collaborative knowledge source: the Encyclopedia Britannica and another collaborative source that is equally well-known: Wikipedia. One of the most notable differences lies in the confidence inspired by the information provided in the Encyclopedia Britannica compared to that provided in Wikipedia, as a result of the reputation of those who develop it and of the reliability of the sources. That is a very positive part of the information in traditional education, strongly based on reputation. Due to its actual philosophy, Wikipedia lacks a reputation mechanism comparable to traditional means of knowledge, however, it is capable of providing almost immediate updates, varied points of view, very useful additional information, links to alternative knowledge sources and in general all the benefits provided by immediate and collaborative information.

A conciliatory and balanced approach between the two extremes could be a version of Wikipedia where all the information has its place but is linked to a reputation ‘index’ of each of the articles based on the level of knowledge of those who write and/or endorse them. At present, Wikipedia can be considered as a sub-set of the proposed Wikipedia where everyone has the same reputation index; based on these ideas, the CF that we will develop will be based on the possibility of awarding greater reputation ‘indexes’ to users with greater knowledge in subjects on which each e-learning RS is based. In this sense, the existing RS can be considered as a sub-set of the ones we propose.

One of the ideas underlined in the philosophy of the actuation of the RS is based on the equality between its users, not only on their possibilities of access to the service, but also above all with regards to the contribution by each one of them to the recommendations that the rest could receive. The usual RS generate the recommendations for each user based on the ratios supplied by the users with contributions most similar to them.

The equal treatment between users is adequate and convenient in the majority of the RS, for example, there is no reason, in advance, to believe that one user is more qualified than another to offer recommendations about movies, journeys, blogs, etc. However, there exists a group of RS in which this situation does not make so much sense. The RS of e-learning are the most paradigmatic in this asymmetric situation; in these RS one can easily make a distinction between advanced and novice users, for example between the ratios generated by teachers and those generated by students, or those supplied by advanced students (for example final years) and those who are starting their studies.

The CF filtering method used can be obtained from the paper [5] in which it is developed the fundamental underlying principal in the CF method proposed together with the equations that uphold it. In the following sections of this paper we explain the design of experiments carried out, we list the results obtained and we explain the conclusions we have reached on applying this CF method to an e-learning RS currently in operation.

### 3. RECOMMENDER SYSTEM EXPERIMENT DESIGN

In order to test the CF method proposed we have used an e-learning RS that has been in operation for 2 years. In this period of time, ratings have been received from 243 students on 224 aspects (44 written documentations of teaching units + 44 proposed exercises + 44 knowledge validation tests + 88 external information sources + 4 lecturers), all of which were in 9 training courses with the subjects: “communications with Java language” and “development of web applications with Java”.

The external information sources (books, on-line documents, websites, etc.) offer 2 references of complementary information to the actual course documentation in each of the teaching units.

The students’ ratings have been collected throughout each course to gauge their opinion, in a range from 0 to 5, on the suitability of each
of the 224 aspects rated as a means of assimilating the knowledge proposed in each of the teaching units or throughout the whole course in the case of rating the lecturers.

The students’ profile was qualified engineers or students in the final year of engineering with a good knowledge of programming in languages other than Java. The technical level of the students in one of the courses (communications) on average was higher than that of the students of the second course. No students have taken part in both courses. The maximum time period permitted to make a rating on a teaching unit was one week after it had finished. The rating of the lecturers was carried out at the end of the course.

Although participation in the RS was not obligatory, due to the interest shown by the lecturers and the students themselves, the end result has led to a much higher ratings percentage than can be found in any public RS database, which has made it necessary to gauge the impact of working with a ratings matrix R that is not very sparse [29,6].

From the students’ point of view, our e-learning RS allowed them to receive personalized recommendations on the suitability of the documentation sources for the classes, which meant they had an idea about which of the 3 documents for each class would best adjust to their preferences. In the first course, the students had recommendations based on the ratings of their classmates and of the lecturers. In the other courses, the recommendation was also processed considering the ratings of students from previous courses.

The matrix R of ratios consists of the 224 elements rated. The matrix C of scores consist of a number of elements $T=44$ and is formed by the marks obtained by each student in the 44 exercises proposed plus the 44 knowledge validation tests, all of which corresponds to the 44 teaching units. As new courses were held, the number of rows in matrices R and C gradually increased to finally reach the total of 243 students participating in this RS.

Once the courses had ended and matrices R and C were complete, we went on to assess the classic factor of accuracy of the RS by calculating the Mean Absolute Error (MAE) and the measure of dispersion of that mean figure: the percentage of perfect predictions [30], by using Pearson Correlation adapted to the CF. After this, we tested the results using the equations designed to be used in the e-learning RS.

4. RESULTS

Figure 1 shows the results of the Mean Absolute Error (MAE) obtained by applying the traditional CF techniques to the data of our e-learning RS (black graph with squares) and the equivalent results using values of $\alpha$ 0.3, 0.5 and 0.7.

The parameter $\alpha$ (with the range of 0 to 1) serves to adjust the weight that we give to the values rated by each user with regards to the weight that we give the evaluations by the users with better scores. Expressed in another way, the parameter $(1-\alpha)$ allows variation of the importance given to the evaluations by people with greater knowledge (scores).

The value of $\alpha$ 0.3 considers to a greater extent the recommendations of users who have a better reputation (or knowledge, in the context of this paper), whilst by using the value of $\alpha$ 0.7 the recommendations of the users that have a better reputation lose a great deal of importance; the value of $\alpha$ 0.5 represents a balance between both tendencies.

The study is carried out using k-neighborhood values: 5, 10, 20 and 40, which allows us to emphasize the most representative values for such a small database and also to show the tendency in the results when the number of k-neighborhoods increases.

As we can see in figure 1, with the exception of $k = 5$ & $\alpha = 0.7$, in all the values processed the error (MAE) is reduced by testing the accuracy of the recommendations using different values of $\alpha$. The best results are obtained by balancing ($\alpha = 0.5$) the importance of the users’ reputation in the recommendation process. This result indicates that, by making this specific RS, we achieve the best estimate by weighting by approximately 50% the ratings of the users with the best knowledge.

This result can be explained by reflecting on the necessary balance that should exist between the choice of users (k-neighborhoods) with sufficiently similar ratings to those of the user who will receive the recommendation as regards the choice of users for whom, despite being less similar, the ratings should be more suitable.

In order to illustrate the previous concept, we can take a look at the following example: In line with my ratings on tourist destinations, a traditional RS could recommend Cancun (Mexico) for my beach holiday in September, which fits in with the ratings that I will have made in the RS: beach, hot, nightlife, etc. however, several experts from travel agencies have given this destination a negative rating at the end of the summer due to the high probability of hurricanes in the area. For the RS to make a recommendation almost exclusively in line with the ratings of these experts would make no sense, as I would probably not have the same taste (ratings) as them (they will not be similar to me), but it is a very good idea to take into account (with certain weighting $\alpha$) their ratings to avoid unwise recommendations from users that are very similar to me, who may be less well informed (in this case about the hurricane season).

Figure 2 shows the percentages obtained of perfect predictions. We consider a perfect prediction to be each situation in which the prediction of the value recommended for one user in one e-learning item matches the value rated by that user for that e-learning item. As an example: we would achieve a perfect prediction when the RS is capable of recommending, with a value of 4, written
documentation of a teaching unit that the user has rated with a 4. This process is only carried out to obtain the graphs shown, as the regular recommendation process only recommends items not rated by the users.

The results shown in Figure 2 lead us to establish that, in this RS, it is not advisable to excessively weight the ratings of users with a better reputation ($\alpha = 0.3$), because we lose the capacity to make correct predictions; however, it is very useful to consider the ratings of these users with a better reputation using more balanced values (around $\alpha = 0.6$), with which we manage to improve the results compared to the use of traditional CF techniques (show in black with squares).

Figure 3 shows the percentages obtained of bad predictions. We consider a bad prediction to be each situation in which the prediction of the value recommended to one user in one e-learning item is different by more than 2 points from the value rated by that user for that e-learning item.

The situation shown in Figure 3 is repeated, in the sense that here it is also counterproductive to overrate the opinions of the experts, but is highly advisable to take them into account in a moderate way (around $\alpha = 0.6$). In this way, we manage to reduce the percentage of bad predictions as regards the traditional results of the RS (shown in Figure 3 as “correlation”).

As we can see, the proposed method of CF adapted to e-learning improves the results of traditional CFs. Moreover, we can provide an approximate average value of how much it is improved by percentage. In Figure 1, by observing the numerical values of the MAE we can determine average improvements of around 17%; this percentage is approximately maintained when we take a closer look at Figures 2 and 3, which leads us to establish a generalized improvement of around 17% when we use the proposed CF system on our e-learning RS.

5. CONCLUSIONS

The recommender systems of e-learning allow the possibility of weighting the importance of the ratings that each user generates, depending on their level of knowledge. In order to implement this concept we provide a new measurement for similarity between users $x$ and $y$, which we call importance and we accompany it with the necessary equations in order to validate the results and adjust each e-learning RS by assigning a greater or lesser importance to users with better knowledge.

The results obtained with our e-learning RS have been very positive; we have achieved improvements of around 17% in the levels of accuracy (MAE) and in the percentages of perfect predictions and bad predictions. The empirical results have also shown the advisability of using balanced values in the importance assigned to users with higher levels of knowledge, in such a way that this new source of knowledge is taken into account, but without excessively compromising the similarity of the k-neighborhoods chosen to make each user’s recommendations.

A well defined field of investigation exists in CF when the nature of the recommender systems allows the incorporation of weighting in the importance of each one of the users, and the collaborative systems of e-learning are named to lead the developments in this new field of investigation.

6. REFERENCES


IEEE/WIC/ACM International Conference on Web Intelligence, 199-205. 10.1109/WI.2004.10158.


