ISBN: 978-972-8924-58-4 © 2008 IADIS

ORIEB, A CRS FOR ACADEMIC ORIENTATION USING QUALITATIVE ASSESSMENTS

Emilio J. Castellano, Luis Martínez
Universidad de Jaén
Campus de las Lagunillas s/n, 23021. Jaén

ABSTRACT

Collaborative Recommender Systems (CRS) are very useful tools that help people to select items in a huge search space, based on the idea that people with similar taste of preferences in an topic make similar decisions concerning to that topic. There are many commercial applications that show the utility of these systems. In this contribution we shall introduce OrieB, a CRS working in the Academic Orientation domain in order to support advisors helping students of secondary school to make decisions about their academic future. OrieB will use students' marks as input data in order to suggest their academic possibilities by providing qualitative recommendations based on the fuzzy linguistic approach.

KEYWORDS

Academic Collaborative Recommender System, Qualitative Information.

1. INTRODUCTION

When students reach certain level in their academic journey, they need to make decisions about their academic or professional future. Many countries have created one figure, called advisor, whose main task is to guide students when they face up these decision situations. But advisors have to manage many students and often it is hard for them to give optimum advices because there are many variables concerning the orientation task, including students' expedient. We think that the use of Collaborative Filtering (CF) using the marks obtained by the students to generate groups of similar students and compute recommendations based on those elections made by previous students could be very useful for academic orientation.

Moreover we consider that marks represent more than simple and single quantifiers. Marks are given by teachers (experts) and set up them taking into account a set of features, skills, knowledge, etc. that a student have in any specific subject; besides, marks reflect also information about student preferences. It is known that whenever a student likes a subject, she usually does her best in it.

Any educational system presents the following features:

- There are students that course subjects.
- There are usually several kinds of subjects: core subjects, elective subjects, and vocational subjects.
- There may be vocational programs, built with groups of subjects having same professional scope.
- Applying an evaluation protocol to the pair subject-student teachers obtain a single value called mark.

Taking into account these assertions a Collaborative Recommender System (CRS) (Schafer et al., 2001, Pazzani, 1999) for Academic Orientation computes recommendations by predicting marks using Collaborative Filtering (CF) and analyzing those predictions by grouping them based on subject topics, programs, grades, etc., could be a valuable system in order to support academic orientation aims.

Following the idea of CRS, we studied its performance in the student guidance domain, in their subject and academic profile choices during their different stages at high/secondary school and University (Castellano et al., 2007). Then we will present OrieB, a CRS which uses students' marks to generate a profile in order to classify them in different groups of similar students concerning their skills and tastes that will be utilized to build recommendations that support advisors guiding students in their academic decisions.

Initially the recommendations computed by OrieB were numerical values, however in fact the aim we chase in this system it is a value to support advisors in their orientation duties, then the use of precise numerical values can be misunderstanding. Therefore due to the existence of uncertainty in the information regarding our objectives the use of approximate values seem adequate to support students decisions and easier to understand by them. We propose the use of qualitative assessments to generate advices that model more intuitively such values than precise numbers. To do so we shall use the fuzzy linguistic approach (Zadeh, 1975) for representing such values and develop a latest version of OrieB that includes an interface with linguistic labels similar to those ones used by advisors in Secondary Schools and High Schools.

The paper is organized as follows. Section 2 makes a brief review of collaborative filtering recommender systems and fuzzy linguistic approach. Section 3 introduces the CRS OrieB's that supports students' decisions in academic orientation. Finally section 4 points out conclusions and future work.

2. RELATED WORK

2.1 Collaborative Recommender Systems

CRS are based on the CF method (Herlocker et al., 1999, Adomavicius and Tuzhilin, 2005) and has been applied to different areas in the literature with successful results. CRS work by collecting human judgments (ratings) for items in a given domain and grouping people who share the same needs, preferences, etc. (Herlocker et al., 1999). Through CF algorithms, a CRS tries to predict the rating that a user would have for an unrated item based on previous ratings of users which have rated similar items to the active user. This method provides several advantages over other techniques used in recommender systems (Herlocker et al., 1999, Pazzani, 1999):

1. Support for filtering items whose content is hard to analyze automatically.
2. Ability to filter items based on quality and taste.
3. Ability to provide serendipitous recommendations (Herlocker et al., 2004).

Although there are several CF approaches they all fulfill three general tasks to perform recommendations demanded by users: analyzing and selecting data sets (Herlocker et al., 1999), grouping users (k-nearest neighbor selection is the most used) (Herlocker et al., 1999, Breese et al., 1998, Adomavicius and Tuzhilin, 2005) and generating predictions based on one group of similar users (Herlocker et al., 1999, Breese et al., 1998, Adomavicius and Tuzhilin, 2005, Pazzani, 1999).

2.2 Fuzzy Linguistic Approach

Usually CRS works in a quantitative setting, with information expressed by means of numerical values. However, the aspects we deal with in our system cannot be appropriate to assess in a quantitative form, but rather in a qualitative one. In that case, a better approach is to use linguistic assessments instead of numerical values. The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables (Martínez et al., 2007, Zadeh, 1975) and has been applied successfully in different areas.

In this approach must be chosen the appropriate linguistic descriptors for the term set and their semantics. In order to do so, an important aspect to analyze is the "granularity of uncertainty", i.e., the level of discrimination among different counts of uncertainty. The universe of the discourse over which the term set is defined can be arbitrary, in this paper we shall use linguistic term sets in the interval [0, 10]. Typical values of cardinality are odd ones, such as 5 or 7.

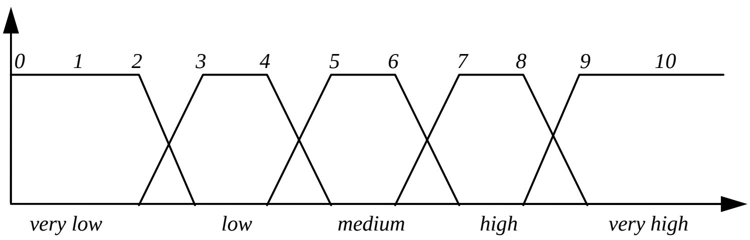

Figure 1. Semantic terms and membership functions.

We generate the linguistic term set by supplying all terms distributed on a scale with a defined total order:

$S = \{s_0 : \textit{very low}; s_1 : \textit{low}; s_2 : \textit{medium}; s_3 : \textit{high}; s_4 : \textit{very high}\}$

The semantics of the terms is given by fuzzy numbers. A computationally efficient way to characterize a fuzzy number is to use a representation based on parameters of its membership function. The linguistic assessments given by the users are just approximate ones; we consider that linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments. The parametric representation is achieved by the 4-tuple (a; b; d; c), where b and d indicate the interval in which the membership value is 1, with a and c indicating the left and right limits of the definition domain of the trapezoidal membership function (see Fig. 1).

3. ORIEB: A WEB BASED CRS FOR ACADEMIC ORIENTATION

OrieB is a CRS using a real dataset collected from several Spanish high schools to perform recommendations about which vocational program and subjects choose in Spanish Bachelor. The aim is to support advisors (not teachers) helping students in this task which, due to the high number of students, becomes harder everyday.

OrieB uses our dataset of students, subjects and marks. The details about algorithms used and results can be seen in (Castellano et al., 2007). The system will support advisors to know which vocational program or programs will fit better to a specific student, which subjects in each vocational and elective subjects will be better to course and finally which core subjects can manifest extra difficulty to a student.

3.1 OrieB's Interface

To obtain recommendations OrieB allows advisors to provide the student's marks which must be represented by means of numbers, as they are assigned to students by teachers. The more marks filled, the more accurate and personalized will be the recommendations obtained.

Also is possible to provide an student ID if the expedient of the student is included in the dataset, so that it would be not necessary to fill any marks as they are already known by the system.

3.2 Recommendations in OrieB

The system computes the predictions for all possible subjects that a student can have in her next grade. Then the system is ready to perform its recommendations. There are 3 types of recommendations coming from the items recommended and the manner they are shown:

3.2.1 Vocational Program Recommendations

The vocational program recommendation will show an ordered list with the vocational programs available. This list is ordered based on predictions computed taking into account the following assertions:

a) One vocational program will be better for a student if marks predicted for its subjects are higher than those of other programs.
b) Probably we won't obtain predictions for all subjects for a specific program, so it isn't enough take into account only marks because they can be not enough. This way, we need also to take into account the number of subjects for the specific vocational program for which predictions were obtained.
c) Finally, we use the *variance* in order to obtain a trust measure, such that, we trust only in those programs that obtain homogeneous predictions for their subjects.

Table 1: Examples of vocational program selection.

	Vocational Program A				Vocational Program B			
	A1	**A2**	**A3**	**A4**	**B1**	**B2**	**B3**	**B4**
Case 1	10	8	7	10	3	5	2	4
Case 2	?	10	?	?	6	?	7	8
Case 3	10	1	2	8	6	7	8	7

As example, we can take the predictions for this two vocational programs A and B, with subjects Ai and Bi, in these three cases.

For case 1 we can check the veracity of assertion (a) because it is clear that the program recommended would be B. With case 2 we confirm (b), because it is more reliable to recommend vocational program B, because the number of predictions we have is greater. And finally we can notice in the case 3 how important is to take into account the variance as said in (c); it couldn't be appropriate to recommend Vocational Program A in this case.

Based on these assertions the system shows to advisor an ordered list with the vocational programs available and a number which express the appropriateness computed by the system, as we can se in Figure 2:

Interest	Vocational program
64.22%	Arts
60%	Humanities and Social Sciences
57%	Natural and Health Sciences
54.5%	Technology

Figure 2. Gathering student's marks.

3.2.2 Subject Recommendation

For specific subject recommendations we group them into two types that are shown separately: vocational program subjects and elective subjects

However, although they belong to distinct groups of subjects, the recommendation is obtained similarly. Vocational program subjects are grouped into each vocational program and then ordered according to their prediction while elective subjects are shown all in the same group and also ordered by their prediction.

The appropriateness of each subject is represented by a linguistic label computed basing on predictions obtained by the system as we will show in section 3.3.

3.2.3 Core Subject Difficulty Advising

If the system computes an unsuccessful prediction for a core subject, it will warn the advisor that probably student will find difficulties coursing it. We consider that system must warn when it obtains predictions lower than 5 (the mean value). A more detailed explanation can be seen in (Castellano et al., 2007).

As shown in Figure 2, a previous implementation of OrieB used numeric percents to express how adequate is a vocational program to the specific student, a subject or for core subject difficulty advising. Due to the qualitative character of this information, we have decided to improve this interface in order to better express recommendations by using linguistic labels, making them more understandable.

3.3 Linguistic Recommendations

Due to the fact, we want to support advisors that suggest students their academic orientation, we think it is not necessary to provide advisors precise numerical values about the suitability of subjects for students, rather than linguistic values seems enough. So, OrieB will provide the recommendations using linguistic labels belonging to the term set shown in Figure 1.

Recommendation	Elective subject
Very high	Mass Media
High	Psichology
High	Computer science
Medium	French (2nd Language)

Figure 3. Subject recommendation in OrieB

When OrieB computes predictions for subjects automatically assign a linguistic label based on the numeric value obtained following the membership function showed in Figure 1. This way, unsuitable subjects will obtain *low* and *very low* values for their recommendations, and better terms correspond with higher mark predictions. Then, the system computes linguistic recommendation based on the membership degree of the numerical prediction obtained by the CF algorithm (see Figure 3).

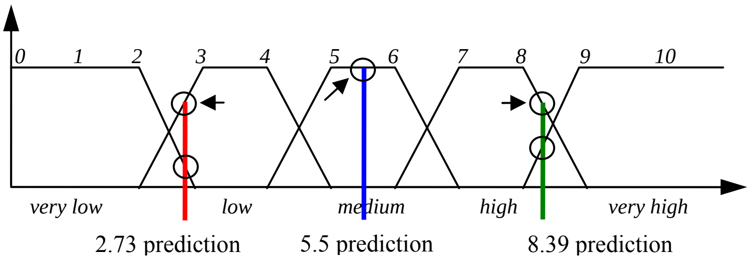

Figure 4. Example of linguistic labels assignment

In Figure 4 we can see three prediction values: 2.73, 5.5 and 8.39. For the label election, OrieB chooses that label which returns a greater belonging value for the membership function. These values belong to the [0,1] interval and correspond to the cut value represented in the graphic. As we can see in figure, first and last predictions have cut values for two labels each, but OrieB will choose the greater value: *low* label for a 2.73 prediction, and *high* label for an 8.39 prediction. Prediction 5.5 only has one cut value and the label assigned is *medium*.

4. CONCLUSIONS

In this contribution we have introduced OrieB, a qualitative web based CRS which offer advisors support to help students in the academic orientation task, giving them linguistic recommendations concerning to academic orientation. The use of linguistic recommendation improves their understanding and fits better the objectives of the system. This system is suitable to any academic level by collecting the proper amount of marks for the target subjects and we are working to include other sources of information.

In the oral presentation will be introduced widely the working of OrieB.

ACKNOWLEDGEMENTS

This contribution is partially financed by research project TIN 2006-02121

REFERENCES

ADOMAVICIUS, G. & TUZHILIN, A. (2005) Toward 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.

BREESE, J. S., HECHERMAN, D. & KADIE, C. (1998) Empirical Analysis of Predictive Algorithms for Collaborative Filtering. IN PUBLISHERS, M. K. (Ed. *Uncertainty in Artificial Intelligence. Proceedings of the Fourteenth Conference.*

CASTELLANO, E. J., MARTÍNEZ, L., BARRANCO, M. J. & PÉREZ, L. G. (2007) Recomendación de Perfiles Académicos Mediante Algoritmos Colaborativos Basados en el Expediente. IN PRESS, I. (Ed. *Conferencia IADIS Ibero-Americana WWW/Internet 2007.* Vilareal (Portugal).

HERLOCKER, J. L., KONSTAN, J. A., BORCHERS, A. & RIEDL, J. (1999) An algorithmic framework for performing collaborative filtering. IN HEARST, M., GEY, F. & TONG, R. (Eds.) *Sigir'99: Proceedings of 22nd International Conference on Research and Development in Information Retrieval.* New York, Assoc Computing Machinery.

HERLOCKER, J. L., KONSTAN, J. A., TERVEEN, K. & RIEDL, J. T. (2004) Evaluating collaborative filtering recommender systems. *Acm Transactions on Information Systems,* 22**,** 5-53.

MARTÍNEZ, L., PÉREZ, L. G. & BARRANCO, M. (2007) A Multi-granular Linguistic Content-Based Recommendation Model. *International Journal of Intelligent Systems,* 22**,** 419-434.

PAZZANI, M. J. (1999) A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review,* 13**,** 393-408.

SCHAFER, J. B., KONSTAN, J. A. & RIEDL, J. (2001) *E-Commerce Recommender Applications*, Kluwer Academic Publishers.

ZADEH, L. A. (1975) The concept of a linguistic variable and its application to aproximate reasoning. *Information Science,* 8 and 9.