ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

An Efficient Recommender System for Predicting Study Track to Students Using Data Mining Techniques

S.Venkata Krishna Kumar[1], S.Padmapriya[2]

Associate Professor, Department of Computer Science, PSG College of Arts and Science, Coimbatore, India [1]

Research Scholar, Department of Computer Science, PSG College of Arts and Science, Coimbatore [2]

Abstract: University or College admission is an intricate decision process but it is an important responsibility of the students to select the correct study track to succeed in their academic life. The complex issue is to assign students to the right academic field when they arrive at the end of basic education stage. In our research we propose a recommender system to students for their higher education which utilizes classification and clustering for recommending the right academic stream and colleges to the students. Instead of the traditional process the proposed system have many advantage such as high prediction accuracy, processing speed and flexibility etc. The decision tree classifier and fuzzy c-means clustering serves as the core design for the college and course prediction system. The system classifies the student and matches them to the proper study tracks according to their profile.

Keywords: Data Mining, Decision Tree, C4.5, Fuzzy C-Means, Recommender System.

I. INTRODUCTION

Advising students about which course to choose after the basic education is a complex task and the purpose of this system is to guide the students to choose the right course according to their interest as well as their personal and academic profile. It helps the students to achieve their academic goals in the specific stream which is recommended to them using the recommender system. The student's interests and strengths in order to choose from a list of potential courses play a vital role in the recommendation. It is equally important to consider the rules for admission into the course and the colleges. Our system provide the solution for the academic advising for the student who are not able to make decisions about their academic life using the rule based expert system and data mining techniques. The suitable recommendation to the students are made by taking the following factors into consideration such as reputation of the college or university, difficulty of the degree, distance from their home, cost of living, and the academic strength of the student as well as the scores secured by them in their schooling. The traditional process suffers mainly from the inaccurate prediction of matching the right students to the right courses and colleges. Therefore we propose a system that automatically predict and recommend the course and college for the student by take into account of multiple variables like demographic information's, secondary school origin, major, study status, etc. for making decision about their higher education. This paper uses machine learning techniques such as clustering and classification for developing the student advising framework. This acquires information from the student data set which contains the personal and academic information of the student. All the machine learning techniques are used to discover the rules for provide the recommendation to the students for their academic success.

The remaining parts of the paper are organized as follows: Section II presents the basic introduction about the data mining techniques and machine learning techniques used for this process. Section III presents the related work done in educational data mining. Section IV presents the proposed work and how the process of the work is carried out. Section V presents the Conclusion and the Future Scope of this process for the enhancement of student assisting process. Then the results are obtained using the WEKA tool kit which is used for the data mining process.

II. MACHINE LEARNING AND DATA MINING

Machine learning aims at building an intelligent system which will be intelligent enough to determine a decision or calculate output based on new inputs after passing the learning phase and being fed with a set of training data. According to the definition of Tom Mitchell [1]: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". Learning can be a supervised learning where the correct output in the training set is made available. Supervised learning is used to solve regression or classification problems. Supervised as the learning Application of machine learning includes classification and regression. Examples of classification problems include identifying an email as a spam, face recognition and hand writing recognition while regression problems include building a model for a system that can be used to predict the output value of the system for a given input. The other type of learning is unsupervised learning where the exact output is unknown. This type of learning is used typically to solve clustering problems.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

Two of the machine leaning techniques is described below.

A. C4.5

C4.5 is a supervised learning algorithm for producing decision tree, it was proposed by Ross Quinlan as an extension of the basic ID3 algorithm. C4.5 is considered a statistical classifier as it can deal with both continuous and discrete attributed and data set with missing attributes values. The standard C4.5 algorithm is as follows:

1) Read set S of examples described by continuous or discrete attributes.
2) Identify base case.
3) Find the attribute which has the highest informational gain.
4) Divide S into S1, S2, S3... according to the values of Abest.
5) Repeat the steps for S1, S2, S3 etc . . .

B. Fuzzy C-Means Clustering

Clustering techniques are known as unsupervised learning methods which is used to group the data that are most similar into one group contains only the similar data. Fuzzy c-means algorithm tries to group the data items into more than one cluster using the membership value.

Step 1: Select the number of clusters c ($2 \leq c \leq n$), exponential weight μ ($1 < \mu < \infty$), initial partition matrix U^0, and the termination criterion ϵ. Also, set the iteration index l to 0.

Step 2: Calculate the fuzzy cluster centers $\{v_i^l \mid i=1, 2, ..., c\}$ by using U^l.

Step 3: Calculate the new partition matrix U^{l+1} by using $\{v_i^l \mid i=1, 2, ..., c\}$.

Step 4: Calculate the new partition matrix $\Delta = \|U^{l+1} - U^l\| = \max_{ij} |u_{ij}^{l+1} - u_{ij}^l|$. If $\Delta > \epsilon$, then set l = l + 1 and go to step 2. If $\Delta \leq \epsilon$, then stop.

III. RELATED WORK

Many researchers have contributed to the field of data mining in higher education. In this section, the researchers will give an overview on a few representative works. C.Vialardi, J. Bravo, L. Shafti, and A. Ortigosa[2], gave a study on recommendation in higher education using data mining techniques. The main purpose of their study is to show how useful data mining can be in the educational domain in order to discover many kinds of knowledge by applying the graduate student data set on the different educational data mining techniques by using data mining tools to discover classification, clusters, association rule, and outlier detection then gave description to their importance in the education domain. E. Gottlieb [3] applied the integer programming technique on the existing higher education student to guide college admission decision. The main goal of their research is to guide the students in their higher education based on the marks of the students in the previous years. This helps decision makers to manage the number of resources and select the track which is useful to student. This study helps on more attention to facilitate taking the correct action at the suitable time to reduce the failure in the academic approach and improve the student's academic performance.L. Chang [4] apply the data mining techniques to predict the college admission to students who in need of the advise for their higher education. Walid Mohamed Aly, Osama Fathy Hegazy and Heba Mohmmed Nagy Rashad [5], gave a case study from the higher education university. Their study was based on applying the educational data mining techniques on the existing student data set from the database university to discovered cluster, decision tree and neural network to show how to evaluate the performance of students with the usage of these techniques. Brijesh Kumar Bhardwaj and Saurabh Pal[6] applied Baysian classification on the student database from the higher education stage. This study aimed at identifying those students which needed more attention to reduce the drop out ratio and take action at a right time which helped to improve the performance of the students and the instructors. A. H. M. Ragab, A. F. S. Mashat, A. M. Khedra, [8] applied a hybrid procedure that was based on the data mining techniques and rules for admission. The main goal of their research is to predict the study track of the students with their profiles and validate it using the auditing process. The college predictior plays an important role in their prediction process.

IV. PROCESS OF RECOMMENDATION

A. Framework Description

The proposed framework uses both classification and clustering techniques to suggest recommendations for a certain courses and colleges for a student which match with their profile. This framework shall include attributes representing:-

- Student academic level profile
- Past History of the Students who have succeed in their academic
- Student personal profile

B. Data Selection and Transformation

In this step only those attributes which is highly relevant to the classification process is selected for data mining. The pre-processing of data removes the missing values and redundancies which are considered as a noise in the data set.

C. Decision Tree Classification Phase

In this phase, a classification algorithm is applied on the educational dataset to find an efficient classifier. The role of the classifier is to classify the students according to their grade of score for a course. The steps in this phase are as follows:-

- Remove the records which is irrelevant
- Use this training dataset, and apply the classification algorithm with the Mark attribute.
- Record the set of rules for the classification algorithm

D. FCM Clustering Phase

In this phase, a clustering algorithm is applied on the educational dataset to divide student records into a number of Clusters based on marks similarity and compare the new user data to the cluster in which it is more similar

then recommend the course and college which match with the profile of the student.

E. Implementation Tool

WEKA tool kit is widely used for the data mining techniques and machine learning which was developed at the University of Waikato in New Zealand. WEKA have many number of machine learning algorithms and data mining algorithms which is used for classification and clustering, Preprocessing etc. To use the collected data in WEKA it should be converted into (arrf) file format.

F. Discussion of Results

The clustering phase clusters the students according to their marks and we have matched the students profile with the cluster in which it mostly suited. For this the Fuzzy C-Means clustering algorithm is used for the accuracy of clustering. For the accuracy it calculate the membership value and then cluster the data according to that value and then for classification the c4.5 algorithm is used to predict the college and the course which is more suited for the new students who logged in to the system.

TABLE I
TRAINING AND SIMULATION ERROR

FP Rate	**Precision**	**Recall**	**F-Measure**	**TP Rate**
0.243	0.609	0.7	0.651	0.7
0.3	0.824	0.757	0.789	0.757
0.28	0.748	0.737	0.74	0.737

TABLE II
CLASSIFIERS ACCURACY FOR C4.5

Evaluation Criteria	**C4.5**
Kappa Statistic	0.5557
Mean Absolute Error(MAE)	0.1904
Root Mean Squared Error(RMSE)	0.398
Relative Absolute Error(RAE)	52.3877%

V. CONCLUSION

In this proposed work we develop a framework which provides assistance to students for selecting their stream in higher education. We group the students into number of clusters and match their profile with the more relevant cluster using FCM algorithm and C4.5 classification algorithm for classifying the students into the course and the college which matches to them. The advantage of the system is the accuracy of the prediction and speed of the result provided. For the accurate prediction the data mining techniques are used and performance analyzed using the WEKA tool.

REFERENCES

[1]Mitchell. T (1997), "Machine Learning", McGraw Hill, ISBN 0-07-042807-7, p-2.

[2]C.Vialardi, J. Bravo, L. Shafti, and A. Ortigosa, "Recommendation in Higher Education Using Data Mining Techniques", Proceedings of the educational data mining, 2009.

[3]E. Gottlieb, "Using integer programming to guide college admissions decisions: a preliminary report", Journal of Computing Sciences in Colleges, Volume 17, Issue 2, Pages: 271 – 279, 2001.

[4] L. Chang ,"Applying Data Mining to Predict College Admissions Yield", Spring 2008 - College of Education, http://www.ed.psu.edu/educ/eps/ ir-certificate/ downloads-forms/hied-860-syllabus-su11.doc.

[5]Walid Mohamed Aly, Osama Fathy Hegazy and Heba Mohmmed Nagy Rashad, "Automated Student Advisory using Machine Learning", IJCA, Nov 2013.

[6] Brijesh Kumar Bhardwaj and Saurabh Pal, "Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011, pp 136-140.

[7]J.S. Moore, "An expert system approach to graduate school admission decisions and academic performance prediction", ScienceDirect ,Volume 26, Issue 5, October 1998, Pages 659–670, http://www.sciencedirect.com/ science/ article/pii/ S030504-8398000085.

[8] A. H. M. Ragab, A. F. S. Mashat, A. M. Khedra," HRSPCA: Hybrid Recommender System for Predicting College Admission" Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference, PP107-113, Kochi, India, 27-29 Nov. 2012.

[9] Q. Li, and B. M. Kim, "Clustering Approach for Hybrid Recommender System" Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI'03), 2003.