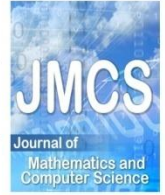




Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com



Recommendations For Academic Major Of Students With Case-Based Reasoning Method To The Case Of Iran

Shiva Asadianfam¹ and Sima Asadianfam²

¹ Department of Computer Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran

² Department of Computer Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran

sh_asadianfam@yahoo.com

sima67_asadianfam@yahoo.com

Article history:

Received June 2014

Accepted July 2014

Available online August 2014

Abstract

In some schools took place in the East Azerbaijan province in this year, a number of students were studied and compared in five different fields of study, based on several criteria, consist of Parental guidance, career, middle and high school grades, courses, etc. The purpose of this study was to predict academic major of students with Case-Based Reasoning method. Factors influencing the choice of the parameters are used in these predictions that the CBR method and other students behavior and process them by Matlab, appropriate field can be suggested for a given test set and measure accurately the amount of the recommendations.

Keywords: Case Based Reasoning Method, Recommender Systems, Data Mining.

1. Introduction

Investigation on education issues, and conformity of results with predefined targets can be effective in showing the problems process of education and action to resolve the problem. What is evident in the present situation, The importance academic major is in high school. The graduates first year of high school based on interest, talent and practice in education and after the assessment and preparedness in the process of changes will be directed to one of the academic major : mathematics and physics, experimental sciences, human sciences, technical and vocational, work and knowledge[1]. The recommender system are known as a sample of decision support systems that the 90s decade as an independent branch opened foot arena research and investigation and help of information about behavior or records of their existing users and their opinions, provide the most appropriate recommendations [2]. Also, Case based reasoning is known as a method of users' behavior have been modeled in dealing with new problems. Thus, the experiences gained in solving problems in the past uses as a guide to solve new problems [3]. In this paper with identifies factors and examine the records of previous students with CBR and by using Matlab,we proposed suitable academic major for the test data sets and measure the accuracy of these recommendations.

This paper structured is follows. In Section 2, we observe the background of the recommender systems and methods of Case-Based Reasoning. In Section 3, apply the method of case based reasoning on the dataset used in several phases. Section 4 contains the results and evaluation obtained from the performance and Section 5 suggests concludes the results of this study.

2. Background

2.1. Recommender System

The recommender systems are used data mining techniques to build and make recommendations using the knowledge Retrieved from users' behavior and features [9]. Simply put, Recommender systems are trying to guess the user's way of thinking (to the assistance of his behavior or his same users and their opinions have) and identify and suggest the most appropriate and closest cases to taste him[2]. The recommender systems generally fall into three main categories.(A) Content-based systems: content information and items information are stored and based on the similarity of items in the history of customer behavior with other items, are provided recommendations to the user. (B) The cumulative filtering (CF): act based on experience from history of customers behavior similar . This system also divided into two categories of user-centric and product-centric. (C) Component Systems: The integration methodologies used in the two systems, content-centric and based on the CF are trying to combine them in different forms, will solve the problems arisen in applying every single one of the two methods above.

2.2. Case-Based Reasoning

Case based reasoning provided a decision support system in solve new problems, this solution is based on the solution similar problems in the past. CBR method has a powerful learning ability, which can use of past experiences, to deal with new problems. CBR system can simple acquire knowledge process by eliminating the time required to extract the solutions of experts. Variable and dynamic status, when an unknown status and solutions are unclear, CBR is an appropriate method of reasoning [3]. In solving problems via this approach, the main activities include:

- Retrieve "the similar samples" to the new issue.
- Use the similar issue response retrieved for preparing a proposed answer to the new issue.
- Revisions of the proposed response , If there is a discrepancy in the new and retrieval issue .
- Keep new samples (the new question and its response) for the future use.

Applying CBR and use of past experiences in reducing the number of iterations required by the algorithm and improved results obtained from the run algorithms was effective and is affordable in terms of time [9].

3. Related Work

There is a wide range of application for the case-based reasoning method , for instance, as mentioned in paper [4], its application in identifying faults in the systems, as seen in the paper [5], in diagnosing the disease, as pointed out in the paper [6], in scheduling, as discussed in the article [7], in making the marketing plans and so on. In addition, in the paper [8], an acceptable number of researches used this method to solve problems in various fields and demonstrated satisfactory results.

4. Materials And Methods

4.1. The Data Set

First, according to interviews advisor and experts in the field identified Effective indexes and was prepared questionnaires to collect students opinions in some schools in East Azarbaijan. Including affect measures can be noted parents guidance grades , advisor guidance, attending college, a job in the future, interest, periods of school, grade school, etc. As shown in Figure 1, at the end of the data collection phase, a text file obtained as follows.

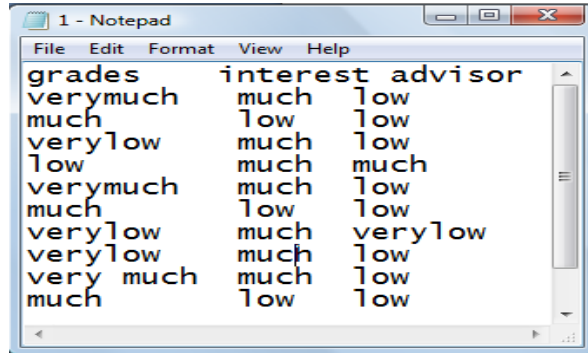


Figure 1 - The dataset used in the text file

4.1.1. Preprocessing Operations

At this stage, some preprocessing operations such as conversions and remove some elements are performed. Matlab software is used to do this. Thus, each of the characters string is converted into a form that can be processed. Also, some factors, such as age was removed due to too little effect. The figure 2 dataset used after pre-processing operations will show in the form of matrix.

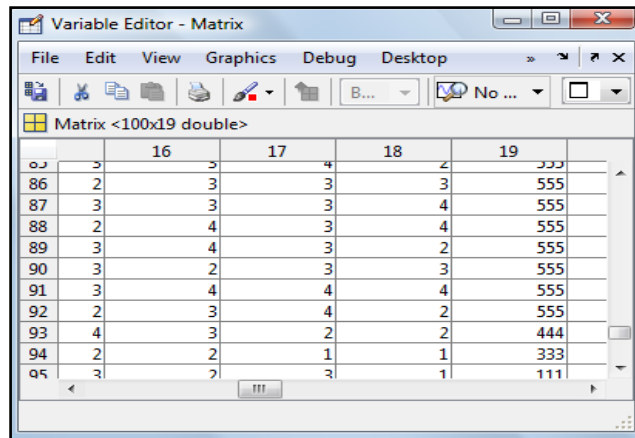


Figure 2 - The dataset used after preprocessing operation

4.2. Pattern Discovery

Once preprocessing operation is done, the data required for statistical applications and data mining methods are ready to patterns discovery. Including data mining methods can be named clustering algorithms, classification algorithms and association rules [10]. In this study, the method of Case

based reasoning is used to recommend the academical major to the students. As was said inference technique based on a prototype is problem-solving method on knowledge-based which acts based on reuse of previous experiences and have emerged from research cognitive science[3]. In this method it is assumed that similar problems have similar solutions. Thus, new problems may be solved using the experience of previous problems. Firstly, when a new problem will be logged, CBR at base samples will be restored that are most similar to it. Secondly, the retrieved sample solutions be reused. Third, the solutions will be reviewed because appropriate with new problem. In the fourth stage, the revised solutions and problem is stored for future reuse.

4.2.2. Calculate The Similarity Of Two Samples Per Case Based Reasoning Method

One of the most complex executive aspects of the CBR method is calculating the degree of similarity of the new problems with previous problem in the recovery phase. In the first stage of CBR, to compute the similarity are needed criterion that assess the similarities to properly. Yet, many criterion have been proposed such as cosine similarity criterion and etc. In this paper, we use K-means clustering for similarities samples. The goal of clustering is that the data divide into several groups. In this division, data of Different groups should be maximum difference than each other. The data contained in a group are very similar to one another. K-means clustering method, despite its simplicity, is the basis for many other clustering methods. This method is an exclusive and flat. For this algorithm has been described many shapes. The simple of this method, first the required number of clusters is randomly chosen points. Considering the closeness degree of data (similarities) attributed to one of these clusters thus new clusters is obtained. By repeating this procedure can be calculated new centers with averaging of data on each repeat and again the data can be attributed to new clusters. This process continues until the change in the data is not obtained [11].

The necessary number to detect clusters of criterion are used sum of squared error (SSE). In other words, the best number of clusters for the division of the sample is required, the SSE criterion is used. The formula for calculating this measure is obtained in (1). With the increasing number of clusters, the rate is reduced by this measure. Therefore, the lowest SSE considered so that increase the number of clusters is no significant impact on the SSE after that. Considering to Figure 3, the number of clusters considered 3, and clustering operation do for detecting the minimum required distance between the sample and the created clusters.

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} \sum_{j=1}^n (m_j^i - x_j)^2 \tag{1}$$

Each sample was compared with the existing portfolio as training data and Finally, a cluster that has the least distance to the test data is selected. The academical major available in the cluster are presented as the solution to the test data and finally, in the maintenance phase, solution revised and the problem is stored for future reuse.

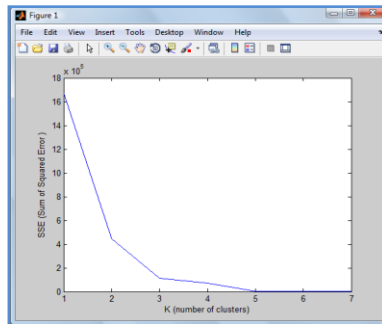


Figure 3 - SSE criterion for determining the number of clusters

4.3. Recommendation

Finally, the goal of this phase is to provide suggestions. K-Fold criterion is used for the selection two sets of training and testing. The original sample is randomly divided into K sub-samples, For each K sub-samples, only one sample obtained as the test data and the remaining K-1 sub-samples are used as training data. This process is repeated a number of K and thus sub-sample is used K as test data exactly . Finally to produce the unit estimate, are averaged between the K results. K-Fold advantage is that the whole set as training data and test data are used with repeated random sub-sample.

The value of parameter k in k-fold method considered equal 6 for preparation of test data and training data and above cases were performed on the samples. Considering to the results, it was determined that the accuracy of proposed system to 68% on the test data.

5. Results

To a fuller understanding of the work done, Precision parameter is used as a criterion to evaluate the work. Precision is a parameter that reflects the accuracy extent of proposed academical major and been matches with the real choice of the students.

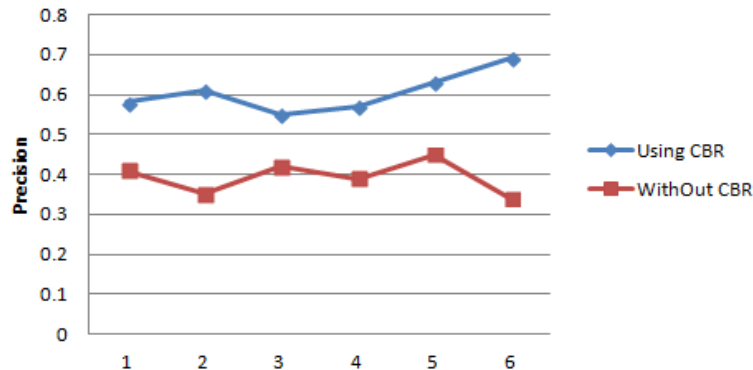


Figure 4 - Chart of evaluate two state prediction using CBR method and without it.

The results obtained comes on using the CBR method until multi-stage of test and evaluation , Precision evaluation factor in the second case (case based reasoning techniques) has high values. Figure 4 depicts results and plots obtained from case based reasoning method and without this method.

6. Conclusions

In recent years, CBR method is shown high capabilities in the various sectors such as decision making, prediction, fault diagnosis, planning, quality control, decision support, distributed computing and data recovery . According to the characteristics of the recommendation field selection can be considered as one of the applications that makes it possible to take advantage of CBR method. In this paper, the results obtained by combining data mining and case based reasoning , we reached the

conclusion that the proposed system in 68% of cases, the test data, to offer a true. Also, with the availability of high volumes of data can be boosted correct offer.

References

- [1] Bish, M.R. The impact of teacher pupil control ideology on the development of middle school self-efficacy. unpublished Ed.d dissertation, university of north carolina at charlotte, 2004.
- [2] Dietmar Jannach , Markus Zanker , Alexander Felfernig , Gerhard Friedrich , An Introduction Recommender Systems , 2011.
- [3] A. Aamodt, E. Plaza ; Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press, Vol. 7: 1, pp. 39-59, 1994.
- [4] Varma, A. and Roddy, N. "ICARUS: Design and deployment of a case-based reasoning system for locomotive diagnostics." Engineering Application of Artificial Intelligence, Vol. 12, No. 6, PP. 681-690, 1999.
- [5] Montani, S., Bellazzi, R., Portinale, L., d'Annunzio, G., Fiocchi, S. and Stefanelli, M. "Diabetic patients management exploiting case-based reasoning techniques", Computer Methods and Program in Biomedicine, Vol. 62, No. 3, PP. 205-218, 2000.
- [6] Schmidt, G. "Case-based reasoning for production scheduling." International Journal of Production Economics, Vol. 56-57, PP. 537-546, 1998.
- [7] SIMON C.K. SHIU, SANKAR K. PAL, Case-Based Reasoning: Concepts, Features and Soft Computing, Applied Intelligence 21, 233-238, 2004
- [8] Mullins R. and et al., A Web Based Intelligent Training System for SMEs, The Electronic Journal of e-Learning, Vol.5, No.1, pp. 39-48, 2007.
- [9] Janet L. Kolodner, An Introduction to Case-Based Reasoning, Artificial Intelligence Review 6, 3--34, 1992.
- [10] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan, Web usage mining: discovery and applications of usage patterns from web data, SIGKDD Explor. 1(12), Jan. 2000
- [11] B. Mobasher, H. Dai, M. Tao, "Discovery and evaluation of aggregate usage profiles for web personalization", Data Mining and Knowledge Discovery (6), pp. 61-82, 2002.