CLUSTER BASED ASSOCIATION RULE MINING FOR COURSES RECOMMENDATION SYSTEM

Wael Ahmad AlZoubi

Applied Science Department, Ajloun University College,
Balqa Applied University, Jordan

ABSTRACT

A course recommender system has a great importance in expecting the selection of courses by students in an university, especially for new students who can't easily select the proper elective courses offered for a specific semester. The computer science department in Ajloun University College at Balqa Applied University (BAU) will be taken as a case study. In this paper, an efficient cluster based rule mining algorithm will be used on a course database to describe a courses recommendation system that assist students to choose elective courses based on students already studied these courses or some of them.

KEYWORDS

Collaborative Filtering, Cluster, Association Rules, Recommendation System.

1. INTRODUCTION

One of the most important applications nowadays is the recommendation system [1] that has a great interest in various areas especially in academic and marketable fields. Recommendation systems are usually used to assist users to select items based on users' preferences [2]. The main goal of a recommendation system is to provide advices to the users to select items that match their needs [9].

In general, almost all available recommendation systems have similar procedure [10] to give the user (the student in this study) what he wants. The steps of this procedure are:

a)   Students' (users) data analysis
b)   Useful information mining by applying an association rule's algorithm.
c)   Offering Items (courses) to users (students) according to their needs.

An efficient cluster based rule mining, that is, Cluster-based association rule (**CBAR**) [3], will be used in the proposed recommender system to let users easily find the preferable courses quickly and to reduce the information overload problem [2]. Mostly, the students can easily satisfy their needs by using online recommendation [10].

2. AN OVERVIEW OF ASSOCIATION RULES MINING

Association rules mining looks for strong relationships between data items that occur commonly together. The problem of mining association rules consists of two steps: (i) Finding all frequent itemsets that have support above a predefined support value. The overall performance of mining association rules algorithms is determined by this step.(ii) Generating the association rules by using the frequent itemsets; which is a direct step.

3. CATEGORIES OF RECOMMENDATION SYSTEMS

According to the method adopted to recommend courses to students, figure 1 displays the main categories of recommendation systems:
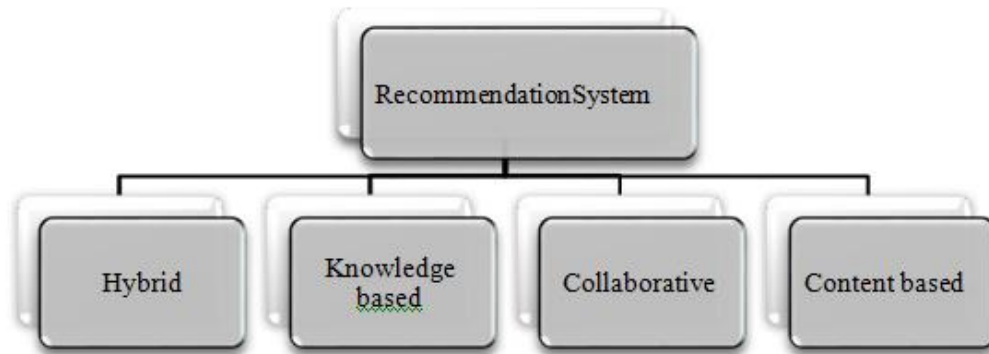

Figure 1. Categories of Recommendation Systems

a) Content-based recommender methods [8]: courses are recommended to a target student based on association between courses' contents and the student profile, ignoring data from other students. This method is not suitable with new students as they haven't previous profile.

b) Collaborative recommender methods [2]: courses are recommended to a target student based on likenesses with other students' preferences. This will be adopted in this research.

c) Knowledge-based recommender methods [3]: inferences are made to conclude a match between a match between students and courses, deep knowledge should be available about the description of courses instead of students' ratings.

d) Hybrid methods [4]: these methods are combination of collaborative recommender methods with content-based or knowledge-based methods.

4. CLUSTER-BASED RULE MINING

In this paper, an efficient algorithm will be used to generate association rules to assist the recommender system suggest proper elective courses to the students. This algorithm is cluster-based association rule (CBAR) [5]. The CBAR method is to create cluster tables by scanning the courses' database only once, and then grouping the student records to the k-th cluster table, where the length of a record is k. Furthermore, the large itemsets are extracted by contrasts with the partial cluster tables. This algorithm has two main advantages:

a) Get rid of considerable amounts of data, and this will reduce the time required to perform datascans.

b) Guarantees the accuracy of the mined rules.

5. PROBLEM DESCRIPTION

The problem of the proposed courses recommendation system is defined as follows:

Given a courses database, some new students, a minimum support threshold (s), and a minimum confidence threshold. The system looks for the courses association rules that have support greater than s and satisfy the minimum confidence constraint in order to recommend the target student to select courses based on courses association rules generated.

The support (rule) measures the times of occurrence of some course to the overall number of students studied elective courses, while confidence measures the percentage of students studied both parts of a rule to the total number of students, they are computed as in equation 1 and 2, respectively [5]:

$$\text{Support } (A \rightarrow B) = P (A \cup B) \qquad ....1$$
$$\text{Confidence } (A \rightarrow B) = \text{support } (A, B) / \text{support } (B) \ldots\ldots 2$$

The proposed collaborative courses recommendation system tries to recommend courses that other students have already taken previously. Typically, students have the right to select some courses from a list of elective courses, mostly the students take advices from other students who took such courses. In the proposed recommendation system, each student will find automatically the suitable choices according to his state after applying the association rules algorithm on some courses database to get courses association rules.

To get courses associations, the first step is building the courses database, after that an efficient cluster based rule mining algorithm will take place. In the courses database, the rows will represent the students and each student has a unique identifier or number while the columns will represent the elective courses selected, we have in our case study five different elective courses, these courses represented by numbers from 1 to 5. Table 1 displays the some of the elective courses offered by computer science department in Balqa Applied University and their sequence numbers as used in our recommender system.

Table 1. Some Elective Courses and Their Equivalent Numbers

Course	Sequence Number
Information Retrieval Systems	1
Advanced Database Systems	2
Cloud Computing	3
Image Processing	4
Robotics and Automated Systems	5

The contents of the table are the grades of each student in all courses studied previously by that student. Table 2 shows an example of a courses database which contains only 20 students from 384 students taken as a case study.

A single letter grading system will be applied "A, B, C, D, E, and F" to denote courses' grades, where A is considered as full mark while F represents zero or fail. The proposed recommendation system generates rules such that:

a)   Advanced DB (A) ∧ Image Processing (E) → Information Retrieval (B) [Support = 0.7, Confidence = 0.45]

This rule means that 70% of students who got A in the "advanced database" course and at the same time got E in "image processing" will take B in the information retrieval course, and 45% of overall students in the case study studied these three courses previously.

b)   Cloud computing (D) → Robotics (D) [Support = 0.9, Confidence = 0.5]

This rule indicates that 90% of students who take D in the cloud computing course will take also D in robotics course, and 50% of the overall students in the case study studied both of these courses.

Table 2 An Example of Course Database

Student ID	Courses	Grade	Student ID	Courses	Grade
1000	1, 2, 4	B, A, E	1010	1, 3, 5	A, B, F
1001	2	B	1011	3	A
1002	1, 5	C, F	1012	2, 3, 5	A, B, B
1003	1, 2, 4, 5	B, A, E, F	1013	1, 2, 3, 4	B, C, A, D
1004	1, 3	B, A	1014	4	C
1005	1, 4, 5	A, D, F	1015	1, 2, 3	A, A, C
1006	2, 3	A, B	1016	3, 5	B, D
1007	5	A	1017	2, 3, 4, 5	B, C, A, F
1008	1, 2, 4	B, A, E	1018	3, 4	A, D
1009	1, 2, 3, 4	A, A, B, E	1019	2, 3, 4	A, B, D

Students are grouped into clusters according to the number of courses taken by each student (transaction length), from the example in table 2, there are 4 clusters, the first cluster of length = 1 which will be denoted as cluster (1) has four students, cluster (2) has five students, cluster (3) has seven students, and finally cluster (4) has four students. Each cluster is studied as a separate unit to find frequent itemsets and in our case courses, that is, the courses that have support greater than predefined minimum support edge.

6. ALGORITHM DESCRIPTION

In this section, the proposed courses recommendation system will be described. this system will use an efficient clustering algorithm as mentioned in section 3 (CBAR) [5] to generate courses association rules, then suitable courses will be recommended to students. The main steps in the proposed recommendation system are displayed in figure 2:
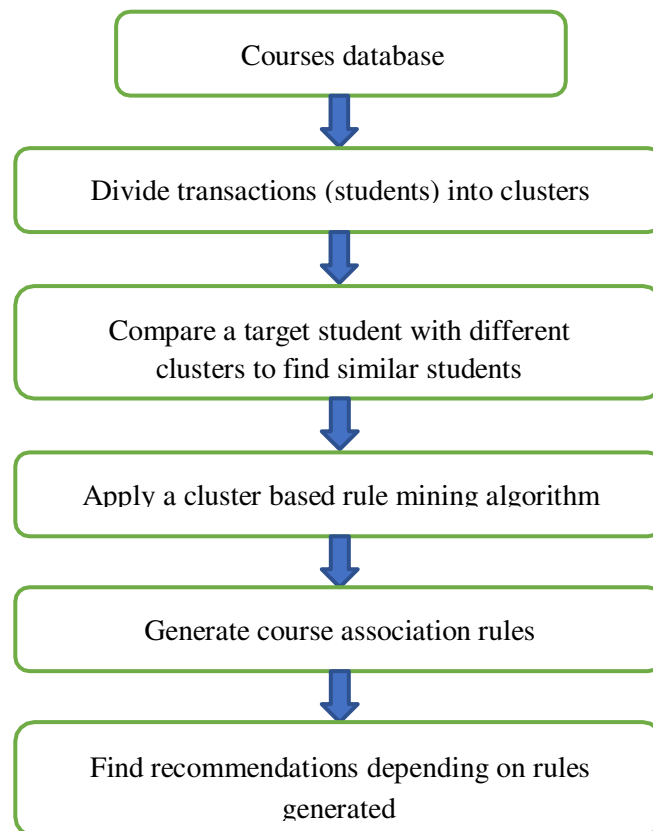
Figure 2.The Proposed Course Recommendation System

a) **Clustering**: This step is necessary to find students who are similar to the target student. The database of courses is scanned only once to create the cluster tables, then the students are grouped to the kth cluster table, where k is the number of courses studied by a student.

b) **Finding similar students**: in this step, the suitable cluster table is only selected from the cluster tables generated in the previous step, then the target student is compared with the previous courses' grades for the mean student of the k-cluster. In this step, the k-nearest neighborhood (k-NN) algorithm [7] is used to select the most n similar groups.

c) **Mining courses associations**: the cluster tables generated in the first step are scanned only once to find frequent set of courses studied together, then the courses association rules will be directly generated.

d) **Recommending courses**: the courses association rules are used to recommend elective courses that the new students prefer. This step is the central step in our recommendation system and was described in details in the next section.

7. RECOMMENDATIONS STRATEGY

The system takes as an input the specified minimum support, the specified minimum confidence, and the course database. For each student, there is one row containing the courses that have been studied by that student and in addition to the grades recorded for that student.

As an output, the system generates courses association rules that satisfy the support and confidence constraints. Then it uses these rules to generate courses recommendations.

The courses recommendation is the main step of the proposed system. The recommendation is performed after courses association rules are generated. The rules generated in step 3 of the previous section are in the form of

[course1:grade1] $\land$ …$\land$ [course$_n$: grade$_n$] $\land$ [course$_n$+1: grade$_n$+1] $\land$ … $\land$ [course$_m$: grade$_m$].

The whole recommendation strategy is described as follows:

1. If the target student has studied the course in the antecedent part of the rule*course1*,…, *coursen* and have obtained *grade1*,…, *graden*, respectively, then the system will recommend the student to take elective courses in the consequent part of the rule *coursen+1*,…, *coursem*

2. The target student may has not taken all the courses in the antecedent part of the rules. In such situation, the rule still be used in the recommendation process and a new constraint called *match* is used to evaluate the quality of the recommendation. The match is defined as the ratio of the number of matched courses between the target student and the antecedent part of the rule to the total courses taken by that student. If *matchrule* is greater than a given threshold, then the rule is used in the recommendation.

3. The target student specifies the minimum accepted grade, so the system recommend courses from the consequent part of the rule if the expected *graden+1*,…, *gradem* are greater than the specified minimum grade.

4. If more than one rule is applicable to a student and therefore many courses are recommended, then the student is advised to choose the rule with the best quality (high support, confidence, and match).

5. The proposed recommendation system gives advices for new students to study courses that have high grades and that have support greater than the predefined minimum support threshold.

Figure 3 shows an example of our recommendation strategy, where the minimum accepted mark is **E** and the minimum match threshold is 50%. Two additional electivecourses, artificial intelligence (AI) and operating system (OS), are taken in the experiments.

Association rule from course database		Support	Confidence
1. [Database: A] $\land$ [Cloud: C] $\land$ [Info. retrieval: B]		50%	50%
2. [Database: A] $\land$ [Image: D] $\land$ [AI:B] $\land$ [Robotics: F]		60%	70%
3. [Info. retrieval: E] $\land$ [OS: C] $\land$ [Cloud: D]		40%	80%
4. [Image: D] $\land$ [Cloud: C] $\land$ [Database:B]		80%	100%
The recommendations using the courses associations are:	Supp	Conf	Match$_\text{rule}$
1. Recommends [Info. retrieval: B]	50%	50%	66%
2. Recommends [AI: B] Don't recommend [Robotics: F] because F<E	60%	70%	66 %
	60%	70%	66%
3. No recommendation because match$_\text{rule}$ < 50%	40%	80%	0%
4. Recommends [Database: B]	80%	100%	66%

Figure 3. Example of the course association rules

8. CONCLUSION

Experiments have shown that association rules mining (ARM) is a required tool for creating recommendations to a set of students. Over the several tests on the course database, we have seen the power of different factors on the performance of the proposed system. The confidence and match of a rule have a great effect on the performance, but a high value of confidence or match may not be the excellent decision. It has been noticed that, by choosing a relatively high value of confidence or matchwe can get better performance.

REFERENCES

[1]   Resnick P. and Varian H., Recommender Systems, Communication of the ACM, 40(3), 56-58, March1997.

[2]   Wiranto, Winarko E., Hartati S., and Wardoyo R., Improving the Prediction Accuracy of Multicriteria Collaborative Filtering by Combination Algorithms, International Journal of Advanced Computer Science and Applications, 5(4), pp. 52 - 58, 2014.

[3]   Philip S., Shola P.B., John A.O., Application of Content-Based Approach in Research Paper Recommendation System for a Digital Library, International Journal of Advanced Computer Science and Applications, 5(10), pp. 37-40, 2014.

[4]   Shih Y. and Liu D., Hybrid Recommendation Approaches: Collaborative Filtering via Valuable Content Information, In the Proceedings of the 38th Annual Hawaii International Conference HICSS '05, 2005, 217-223.

[5]   Tsay, Y.-J. & Chiang, J.-Y. 2005. CBAR: an efficient method for mining association rules. Knowledge-Based Systems 18 (2005), pp. 99–105.

[6]   Neha Aggarwal et al. A Mid – Point based k-mean Clustering Algorithm for Data mining. International Journal on Computer Science and Engineering (IJCSE).Vol. 4 No. 06 June 2012. Pp 1174 – 1180.

[7]   Cheng, D. et al. kNN Algorithm with Data-Driven k Value. International Conference on Advanced Data Mining and Applications. December 2014. DOI: 10.1007/978-3- 319-14717-8_39.

[8]   Balabanovic M, Shoham Y. Fab: content-based, collaborative recommendation. Communications of the ACM; 1997 March; 40(3):66-72.

[9]   AlBadarneh, Amer & Jamal AlSakran, An Automated Recommender System for Course Selection. International Journal of Advaned Computer Science and Applications, 2016.

[10]   Raghad Obeidat, Rehab Duwairi, Ahmad Al-Aiad. A Collaborative Recommendation System for Online Courses Recommendations, 2019. International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep- ML), 2019

AUTHOR

Dr. Wael Ahmad AlZoubi Assistant professor at computer science department, Ajloun University College, Balqa Applied University, Jordan, since February of 2014. My research interests are in image processing, pattern recognition, and natural language processing.