Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 93 (2016) 917 – 923
Procedia Computer Science

6th International Conference On Advances In Computing & Communications, ICACC 2016, 6-8 September 2016, Cochin, India

Application of Predictive Analytics in Intelligent Course Recommendation

Deepthi Upendran[a]*, Shiffon Chatterjee[a], Sindhumol S[a], Kamal Bijlani[a]

[a]*Amrita e-Learning Research Lab (AERL), Amrita School of Engineering, Amritapuri, Amrita Vishwa Vidyapeetham, Amrita University, India.*

Abstract

Students who pursue admission to colleges usually experience a difficulty to select a course. In this paper, we propose a course recommendation system to find out the courses which are apt for a student pursuing admission to the college. Typically, the prediction is based on the career goal or the present job trend. In this system proposed, the prediction is formulated based on the grades acquired by the student in twelfth standard; which is taken as a sign of the previous academic performance and cognitive ability of the student. A model is generated from the legacy data or data from the students who have completed the course successfully. This model is used for predicting the courses for new students. The idea behind this approach is that when a student with specific set of skills is successful in a course then another student with similar set of skills will have a higher success probability in the said course.

Peer-review under responsibility of the Organizing Committee of ICACC 2016

*Keywords:*legacy data; course choice; cognitive ability; decision paralyses; apriori principle.

1. Introduction

Students seeking admission to colleges often face difficulties in terms of choosing courses and academic programmes. They are faced with a wide variety of courses to choose from. At the same time, their knowledge and experience is not adequate to decide which courses would be most suitable for them.

Further, students might be interested to pursue certain careers after obtaining their college degree, but may not be aware of which courses would match with their career goals; also whether they have the requisite skills to pursue

* Corresponding author.
*E-mail address:*deepthi.iota@gmail.com

Peer-review under responsibility of the Organizing Committee of ICACC 2016
doi:10.1016/j.procs.2016.07.267

these courses. Often, the course and academic programme selection is done on the basis of current trends in the job market, advice from family members, and peer influence. A wrong decision at this stage might affect the students' careers adversely. Hence the decision to choose academic courses and programmes is very critical.

This paper proposes a course recommendation system designed to help the student to short-list the courses that suit the grades of the student. The grades of the student can be taken as an indication of the cognitive ability of the student in a particular course. This helps us to determine how well the student might be able to do in an academic programme related to a course or group of courses. While the success of a student in any course is likely to be influenced by different factors including the student's interest and motivation to pursue the course and the career interest of the students; the focus of this paper is on prior academic achievement and cognitive ability of the students as reflected through their grades.

The course recommendation system proposed in this study gives us some suggestion based on a set of rules. This set of rules is developed using the data of previous students who have successfully completed various courses. When a new student's data is given to the system, it searches for the previous students who have the data and also traces out the courses in which the previous students were successful. So, the attributes of the legacy data (data of previous students) and new data are matched thereby predicting the success of the student.

The system provides a list of courses with better success probability. This helps to reduce the confusion of the student as they get a better idea about the courses they have to focus on. Thus, there is an elimination method to ignore the courses that do not suit the student. This reduces the effort in pointing out the courses in which the student is likely to succeed.

2. Related Works

The Degree Compass[1] has been used to recommend courses and curriculum for students in post-secondary degree programmes. This system was developed to point out the course choices that will benefit the student; and also to help to bridge the achievement gap across students from different socio-economic backgrounds. Similar to the work proposed in this paper, the Degree Compass aims to help the student to overcome decision paralyses by reducing the number of choices on the basis of which the student decides which courses to opt for. The Degree Compass predicts the sequences of courses through a programme while proposed system suggests the major programme choices that best fit the student.

The Course Agent[2] is another system which contains details of available courses. This system collects the data from the students regarding their career interest and goals. It is a community based system and information is manipulated based on feedbacks from the user. The two types of feedback are implicit and explicit feedbacks. The implicit feedback is extracted from the user actions and explicit is collected from user directly. The course recommendation system proposed in the current paper, aims to include the career interest of the student as an attribute, as part of future work.

RARE[9] or Recommender System based on Association rule is another system which recommends the courses and the sequence of courses to be taken by the student. It is done based on three modes: Content based filtering, collaborative filtering and hybrid filtering. In the content based filtering the past preferences of the user are considered, while in collaborative filtering the past preferences of other users with similar taste are shared.

Similar to the proposed system, RARE also faces cold start problem, i.e., it requires some initial data to set up the system and this takes some times. Unless it has some data to find out the pattern it will not be able to give recommendations effectively. RARE as well as the proposed system uses association rule to figure out the recommendations. The RARE proposes the course sequence while proposed system points out the major programme.

3. Solution Approach

3.1. System Architecture of the Proposed Model

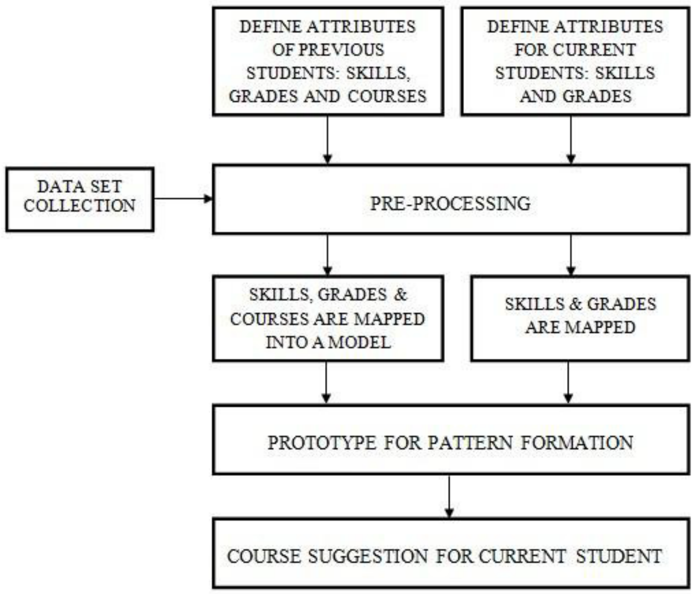

Fig. 1. Block diagram of the proposed model

The architecture of the proposed system sorts out and defines the attributes required to analyze the skills[3] of the student. The skills of both previous students(legacy data) and new students have to be analyzed. The next task is to collect the data with the specified attribute.

The attributes are defined on the basis of grades in twelfth standard for various subjects taken as a sign of the student's prior academic achievement and cognitive ability with respect to each subject. When the legacy data (previous student data) is ready, the model has to be generated using the data.

The proposed model helps us to map the skills of the previous students to the courses in which they have succeeded. This model will be used for predicting courses in which the new students are likely to succeed. While building the model, students' skills and their success in the course act as the input. Success of the new student is predicted with the help of skill map specified in the model. In order to achieve that, we use skill sets from new students as input to the model.

The proposed model will predict the output, i.e., whether the student will succeed in a given course. On the basis of this prediction, we can determine the courses in which the student will probably succeed.

3.2. Defining Attributes

Table 1. The attributes of previous student data set (legacy data)

Attribute	Description
Grades	12^{th} Standard grades of the student
Success/failure	determines whether the student has succeeded or failed in the given course

Table 1, describes the attributes of the system based on the legacy data which shows the grades of the student and their final results in the respective courses. At the twelfth grade, the student will be having various subjects to study and they are Mathematics, Biology, Physics, Chemistry, English etc. The result can be success or failure.

Table 2. The attributes of new student data set

Attribute	Description
Grades	12^{th} Standard grades of the student

In Table 2, the attribute set contains the grades of the new students for whom we have to predict the result.

3.3. Predictive Analytics

Predictive analytics is the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. Predictive models use known results to develop (or train) a model that can be used to predict values for different or new data. The modelling results in predictions that represents a probability of the target variable (for example, in our system the target variable is the success of the student in various courses) based on estimated significance from a set of input variables. This is different from descriptive models that help to understand what happened or diagnostic models that help to understand key relationships and determine why something happened[10].

"Predictor"[7] is the central element of predictive analytics which is the variable used to predict the future behaviour. It uses statistical algorithms and machine learning techniques to identify the probability of future output. The predictive modelling uses three or more predictive variables for implementing the model.

3.4. Association Rule

Association rules[5] are if/then statements that help to uncover relationships between seemingly unrelated data in a relational database or other information repository. An association rule has two parts, an antecedent (if) and a consequent (then). Antecedent is an item found in the data. Consequent is an item that is found in combination with the antecedent.

Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true.

Here we apply market basket analysis[4], it uses association rules for analyzing and predicting customer behaviour. The market basket analysis is the study of items that are purchased or grouped together in a single transaction or multiple, sequential transactions[8]. They play an important part in shopping basket data analysis, product clustering, catalogue design and store layout. An association rule is an implication expression of the form $X \rightarrow Y$, where X and Y are disjoint item sets, i.e., $X \cap Y = \emptyset$ The strength of an association rule can be measured in terms of its support and confidence.

Support determines how often a rule is applicable to a given data set, while confidence determines how frequently items in Y appear in transactions that contain X. The formal definitions of these metrics are:

$$Support, s = \frac{\sigma\ (X \cup Y)}{N} \tag{1}$$

$$Confidence, c = \frac{\sigma\ (X \cup Y)}{\sigma\ Y} \tag{2}$$

where σ stands for probability.

3.5. Apriori Principle

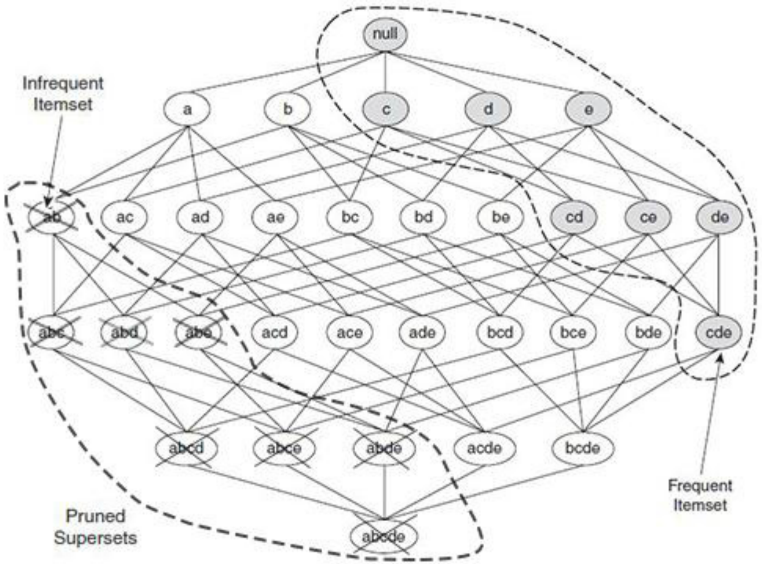

Fig. 2. Apriori principle interms of frequent itemsets and infrequent itemsets (Anti-monotone property)

Apriori principle[5] reduces the number of candidate itemsets obtained during frequent itemset generation based on the support value. The use of support measure for reducing candidate itemsets is given by the following principle.
Theorem: *If an itemset is frequent, then all of its subsets must also be frequent*.[5]

The concept of Apriori principle can be better explained using Fig. 2:

Consider c, d, e as a frequent itemset. If a transaction contains c, d, e then it must also contain its subsets, c; d; e; c,d; c,e; d,e; c,d,e. Hence, if c,d,e is frequent, then all its subsets will also be frequent. Conversely, if an itemset a, b is infrequent then, its supersets are also infrequent. In the Fig. 3, the entire sub-graph that contains the supersets of a, b is pruned as soon as a, b found as infrequent. When the exponential search space is reduced based on the support measure it is called as **support-based pruning**[5]. Thus a pruning strategy is applied by an important feature of support measure, i.e , the anti-moonotone property which states that the support of an itemset will never exceed the support of its subsets.

4. Result and Analysis

As the first step legacy data, i.e., the data from the previous students has to be collected along with their success or failure in various courses. It includes the twelfth standard subject-wise grades of a group of students who have done the selected course, here we have considered Mathematics course. We also have to get the data of the new students for whom we are supposed to suggest the courses. The data is categorised into various attributes based on the marks and the cognitive skills. The marks of the core subjects such as Mathematics, Physics, Chemistry, Biology, Computer Science and English are considered as attributes.

In the current work, we have considered only the marks of previous students. The data of 1000 students was generated synthetically using random number generating algorithm. By using this test data, a set of rules were generated.

Table 3. Synthetically generated mark-list for Mathematics Course

Biology	Computer Science	Chemistry	English	Maths	Physics	Success in course
64	82	81	78	69	91	FALSE
94	76	63	62	68	65	FALSE
88	64	72	75	85	63	FALSE
89	89	96	74	89	87	TRUE
86	85	93	72	93	79	TRUE
81	91	60	93	61	80	FALSE
73	83	86	78	68	73	FALSE
86	92	92	92	79	73	TRUE
65	83	70	92	75	96	TRUE
66	98	63	71	74	78	FALSE

From this set of rules, we consider only the rules with success as the "consequent". The rule with maximum support and confidence is considered for generating the model.

Another test data of 100 students was generated synthetically using random number generating algorithm and it is generated to test the rules generated. It is tested using the rule determined from the first data set. Now we have to cross check whether the success attribute in the synthetic data matches with the success attribute generated using the rule.

The experiment was conducted on various datasets for different courses. Here, the experiment is explained using the Mathematics course which was selected randomly. The success of the student is determined by taking the weighted average of marks of the student. Table 3 shows a sample of the mark-list. In order to generate a set of rules, Association Rule is applied to this mark-list. Thus, it generates a set of rules. Among these rules, the one which has success as the consequent as well as with maximum support and confidence will be considered for the course.

Table 4 shows the list of rules generated and the corresponding mapping of attribute number and courses. The rule selected indicates the relation between the subjects in twelfth standard as well as the success in a particular course. Here, the grades of Mathematics and Physics help to predict the success of the student in the Mathematics course. We observe that student with good scores in those set of subjects based on the rule has a higher probability of success in the said course. This does not mean that the other students have less success or they have failure probability in the said course. We only make a prediction of chances to succeed in a course.

The rules are tested using a sample data set of 100 students. The result obtained indicates that the system is 90% efficient.

A student who is not that good at Mathematics and Physics can also do well in a Mathematics course, if he or she works hard towards it. The success is dependent on many other factors such as interest of the student, career goals, and the skill in grasping the subject, etc.

Table 4. Rule selected from the list of rules generated (1:Biology; 2:Computer Science; 3: Chemistry; 4: English; 5:Maths; 6:Physics; 7:Success in course.)

Antecedent	Consequent	Support	Confidence	Lift
[1,5]	7	0.189	0.759036145	1.552221155
[2,5]	7	0.198	0.798387097	1.632693449
[3,6]	7	0.204	0.784615385	1.604530439
[3,5]	7	0.226	0.926229508	1.894129874
4	7	0.243	0.486	0.993865031
[5,6]	7	0.246	0.991935484	2.028497922
1	7	0.248	0.496	1.014314928
2	7	0.253	0.506	1.034764826
3	7	0.309	0.618	1.263803681

It is observed that the students will have more success probability in courses suggested using the rules generated, as the system eliminates the courses with less success probability. This system is meant to reduce the effort of the student to choose a course from large list of courses.

5. Conclusion

In this paper, a course recommendation system was proposed to suggest courses that suit the cognitive abilities and prior academic achievements of the student as reflected through their grades. In the process, courses in which the student has lower success probability are eliminated. By facilitating better-informed course choice, the system helps to increase the probability of success of the student in a course. Thus, the system is likely to benefit course choice, course performance, as well as career options open to the student in future.

However, there are some aspects which need to be considered while interpreting the recommendations from the system. The system suggests courses with higher success probability. However, this does not mean that if students select courses other than the shortlisted ones then they may not succeed. Success is likely to depend on a number of other factors, including interest, motivation, and effort. For instance, a student with high levels of interest and perseverance in a course may succeed despite initial disadvantages such as low scores in the qualifying examination.

As part of future work, the current model will be enhanced by considering, in addition to cognitive aspects, certain emotional and affective attributes such as student interest or career goals. Further, an attempt will be made to include measures of cognitive ability beyond the grades obtained by the student. In course of time, a repository on cognitive and affective attributes could be built. With an increase in the number of independent attributes, the course recommendation system could be developed more effectively and efficiently.

References

1. Denley T. How predictive analytics and choice architecture can improve student success. Research & Practice in Assessment. 2014 Dec 1;9.
2. Farzan R, Brusilovsky P. Social navigation support in a course recommendation system. InAdaptive hypermedia and adaptive web-based systems 2006 Jun 21 (pp. 91-100). Springer Berlin Heidelberg.
3. Bier N, Lip S, Strader R, Thille C, Zimmaro D. An Approach to Knowledge Component/Skill Modeling in Online Courses. Open Learning. 2014 Apr 30.
4. Introduction-to-market-basket-analysis, http://blogs.mathworks.com/loren/2015/01/29/introduction-to-market-basket-analysis/dd75d035-2129-4417-b848-b2725411df56.
5. Introduction to Data Mining, http://www.users.cs.umn.edu/ kumar/dmbook/ch6.pdf.
6. Predictive Analytics : What it is and why it matters, http://www.sas.com/en_us/insights/analytics/predictive-analytics.html.
7. Predictive Analytics, http://searchsalesforce.techtarget.com/definition/ predictive-analytics
8. Market Basket Analysis, http://www.statsoft.com/Solutions/Marketing /Market-Basket-Analysis
9. Bendakir N, Aïmeur E. Using association rules for course recommendation. InProceedings of the AAAI Workshop on Educational Data Mining 2006 Jul 16 (Vol. 3).
10. Predicitve Modeling, http://searchdatamanagement.techtarget.com/ definition/predictive-modeling.