BUILDING SUBJECT PRE-REQUISITE COURSE MAPS FOR ON-GOING ACADEMIC COURSES

Anamika Jain
Research Scholar
Bhilai Institute of Technology, Durg
Chhattisgarh, India
anamika20001@rediffmail.com

Dr. Arpana Rawal
Professor
Bhilai Institute of Technology, Durg
Chhattisgarh, India
arpana.rawal@gmail.com

ABSTRACT
Educational data mining is an emerging discipline, concerned with developing methods for exploring the vast data sets that come from live educational settings in learning environment.. It is a continual process, keeping pace of vision and mission of an institution, also addressing to issues in an innovative , ethical and responsive manner to meet the academic and administrative objectives. Invariably EDM studies orient towards a mandatory step of undertaking a cross sectional view of attributes contributing to learning patterns of the students as a whole. In this paper, a novel method is proposed to build subject pre-requisite course maps of higher semesters for students pursuing bachelors of engineering courses. The dependency computations are done by analyzing their performance in subject prerequisites of previous semesters. This piece of work focuses on entirely new kind of feature vector that is heavily subsumed to affect the most critical mining objective that is predicting subject wise students academic performance well before they face their end semester examinations.

CCS Concepts
• Information systems→Database management system engines • Computing methodologies→Massively parallel and high-performance simulations. This is just an example, please use the correct category and subject descriptors for your submission. The ACM Computing Classification Scheme:

http://www.acm.org/about/class/class/2012. Please read the HOW TO CLASSIFY WORKS USING ACM'S COMPUTING CLASSIFICATION SYSTEM for instructions on how to classify your document using the 2012 ACM Computing Classification System and insert the index terms into your Microsoft Word source file.

Keywords
Educational Data Mining, Association Rule Mining, Subject Pre-requisites, Strong Association Rules, Rule Generation

1. INTRODUCTION
Over years, higher education system in India has taken the shape of business oriented enterprise-cum-knowledge imparting portfolio gaining a fierce competitive edge at the end of stakeholders and investors. The education system in our country encounters enormous, diversified & manifold challenges with stretching ends constrained from contemporary curriculum development, quality assurance, accreditation schemes, policy planning and Governance ethics. Nevertheless, gradual but consistent schematic amendments are taking place in order to make the students' put their academic efforts at a regular learning pace. Academic analytics is an upcoming machine learning paradigm that was introduced to suffice such an objective of enhancing higher educational quality standards. One way to achieve quality objectives in higher education system is by discovering analytical knowledge from educational datasets to study the major contributing attributes that affect the students' performance directly or indirectly. If Academic communities are able to identify the weak performers and slow learners much earlier to their duration of examinations using EDM practices, this knowledge can aid them in taking pro-active actions, so as to reframe better educational policies and strategies for enhancing academic performance of these students with their upgraded evaluation settings.

2. ARM IN EDM PRACTICES
A huge span of time has already been spent by EDM researchers revealing students' profile patterns and understanding students' learning behaviour using predictive modeling methods to identify drop out students, all encompassed together in a field of software development called Academic Analytics. The realm combines technology, information retrieval, management of data, statistical analysis to tap these potential patterns that help faculty and advisors to become more proactive in identifying at-risk students and responding to them for counseling activities and subsequent remedial actions accordingly. Relationship mining has historically been the most exploited field of EDM research and remains extremely prominent to this day. Relationship mining is used to find the relationships between the created model's predictions and additional variable. These set of techniques enable researchers to formulate association rules, correlation rules, sequential / temporal association rules and rules of causality. Association Rule Mining (ARM) is one of the most important and well surveyed mining techniques intended to identify strong rules discovered in databases using different measures of interestingness. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories [1] [2] [3].

Towards the satisfaction of teaching learning objective in academia, huge span of time has already been spent analyzing varied students' profile patterns; however very little effort is put up for identifying subject wise learning levels of students by exploring wide spectrum of attributes they exhibit while studying

their ongoing courses (subjects). With an exhaustive survey, a general opinion upon which the entire EDM community is unanimously agreeable that it is always better to undertake rational educational parameters of the present rather than past to construct predictive modeling of various forms [3] [4] [5] [6].

3. PROBLEM HYPOTHESIS

For a given courseware, if the syllabus in any academic curriculum provides a glimpse of the course map defining subject pre-requisites for the subjects in forthcoming semesters, this may increase a concern of interest, inquisitiveness and sincerity among these grooming minds towards the on-going courses. Verbal orientation and briefing sessions can be given on these course maps in early days at the commencement of on-going academic semesters. A set of subject pre-requisites for any subject S is defined as a set: S1,S2,S3……Sk, that the student studies in the curriculum of previous semesters and helpful for concept building and thorough learning of the subsequent subject, 'S' in current ongoing semester. Another way of interpreting the above problem formulation is that, upon analyzing the scoring patterns of passed-out batches in hypothesized set of subject pre-requisites, one can reveal strong co-relations between the subjects of on-going courses and some of the selected pre-requisites called as strong subject pre-requisites. Such computations are performed in the following section using Association rule mining method.

4. DATA SETS AND EXPERIMENTATION

4.1 Data Collection

The students' data sets collected in the current research study pertain to the different courses (subjects) pursued by the students of engineering final and pre-final years. Students' performances in respective subject pre-requisites were collected from the departmental records of result summaries pertaining to three passed out batches of computer science and engineering discipline. In the current study, end semester portion of the subject scores obtained, out of 'eighty' were recorded as separate input attributes to the experimental setup.

In the initial stages of experimentation, the subjects that were perceived to be more or less influential for smooth grasping of a subject in higher semesters of the mentioned discipline were included as participating candidates for hypothesizing their pre-requisite schemas. Table 1 shows the pre-requisite schemas for two subjects of sixth and eighth semesters, hypothesized after taking expert opinions by eminent subject experts, who are also the nominated Board of Studies members in the affiliated University, for the formulation of revised syllabi and schemes for courses of these semesters. The abbreviations used for each subject pre-requisite have their interpretation, the annotations borrowed from the syllabi of almost all Indian Universities.

The experimental setup begins by fetching the 'marks obtained' data columns by the passed-out student instances in the above mentioned subject pre-requisites. Student's scores in these prerequisites various subject pre-requisites of previous semesters were analyzed and converted to nominal values in the form of 0 & 1. The WEKA ARM classifier begins with configuring the tool option using suitably chosen values of 'support' (τ) and 'confidence' (σ) mining parameters.

Table 1 Recommended subject pre-requisites for sixth and eighth ongoing academic semesters in Computer Science and engineering discipline

Subjects under Experiments (Candidate Subjects)	Hypothesized subject pre-requisites
B.E. (Eighth semester)	
Artificial Intelligence (AI)	NA, DIS, CF, CM, OOPS, M3, PSLB, DBMS
B.E. (Sixth semester)	
Computer Graphics (CG)	M3, ADA, DS, CSA, CF, DIS

Owing to fact that in some of the configuration settings, too many or almost nil association rules were extracted, the values of support = (0.3) and confidence = (0.6) were finally selected as optimal values for extraction of interesting association rules for every subject mentioned in the table 1, as a consequence of multiple execution trials with varying combinations of configuration setting combinations. One such set of interesting correlations for B.E. 8th semester subject, 'Artificial Intelligence & Expert Systems' is shown in table 2 along with confidence measures.

Table 2 Strong association rules for Artificial Intelligence & Expert Systems

1	CM=1 OOPS=1 134 ==> AIES=1 97 conf:(0.72)
2	NA=1 CF=1 144 ==> AIES=1 97 conf:(0.67)
3	CF=1 OOPS=1 153 ==> AIES=1 101 conf:(0.66)
4	OOPS=1 184 ==> AIES=1 121 conf:(0.66)
5	M3=1 CF=1 151 ==> AIES=1 99 conf:(0.66)
6	CF=1 CM=1 151 ==> AIES=1 99 conf:(0.66)
7	A=1 182 ==> AIES=1 117 conf:(0.64)
8	PSLB=1 CF=1 154 ==> AIES=1 99 conf:(0.64)
9	CM=1 DIS=1 152 ==> AIES=1 97 conf:(0.64)

Table 3 Dependency rule generation for subject pre-requisites of AIES and CG courses

Candidate Subject	Rule Interpretation using subject pre-requisites
Artificial Intelligence and Expert Systems (AIES) (8^{th} Semester)	IF (NA=1) or (OOPS=1) THEN AI=1 ELSE IF (CM=1 and OOPS=1) or (NA=1 and CF=1) or (CF=1 and OOPS=1) or (M3=1 and CF=1) or (CF=1 and CM=1)or

	(PSLB=1 and CF=1) or (CM=1 and DIS=1) or (CF=1 AND DBMS=1) THEN AI=1 ELSE AI=0

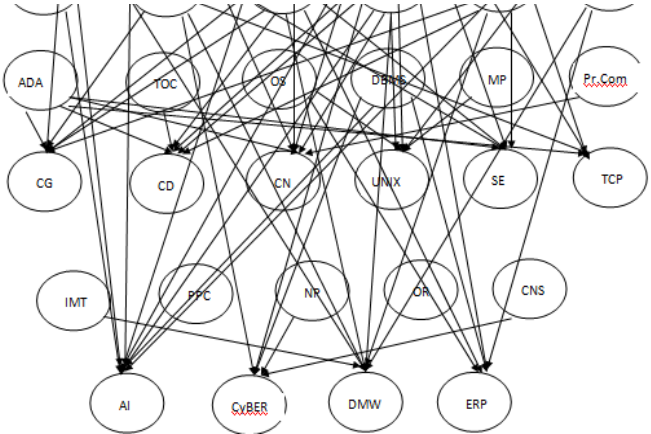
Computer Graphics
(CG)
(6th Semester)
	IF (M3=1 or DS=1 or CF=1 or CSA=1 or ADA=1)
	THEN CG=1
	ELSE
	IF(CSA=1 and ADA=1)or
	(DS=1 and ADA=1)or
	(M3=1 and ADA=1)or
	(M3=1 and DIS=1)or
	(DIS=1 and ADA=1)THEN CG=1
	ELSE CG=0

In the next step of computation, the strong association rules for the target (candidate) subject (with confidence value above threshold) were exploited for interpreted rule interpretation task in the form of IF-THEN statements holding compound conditional expressions as antecedent stubs evaluating to strong dependency of the candidate subject on the specified pre-requisites using following heuristics:

1. Upon analyzing each rule it was observed that some pre-requisites contributed towards the dependency of candidate subject (consequent) independently. For instance, rules 2,4,7,8 and 9 contribute the dependency to candidate subject 'CG' independently and so appear as disjunctive components of the compound conditional expression in the antecedent portion of rule.

2. On the other hand multiple pre-requisites appear as conjunctive components and at least one occurrence out of many such combinations may contribute to strong dependency of the subject candidate; all together combined using disjunction connectives.

Figure 1 Subject pre requisite course map for sixth and eighth semesters



The rule interpretation task for two subjects, each from sixth and eighth semesters is illustrated in table 3.

Subject pre-requisite course maps identified for all semesters in the any academic curriculum is bound to benefit all learners and teachers in teaching-learning environments. The findings of this study conclude that some subjects act as heavily determining criteria towards the prediction of student performance in higher semester subjects. Dependency structure thus obtained in the form of subject pre-requisite course map can be further exploited in prediction of subject wise scores of each student as a mark of progressive evaluation framework, well before they face their final examinations of the on-going semester. Identifying the subject attributes, that contributes the most towards the students' performance help to improve the support services for students who perform poorly in their studies in early stages amidst their academic semester. It may be noted that similar experiments can be extended upon total scores of the subject pre-requisites, following the cumulative marking scheme based on internal assessments too.

5. ACKNOWLEDGMENTS

This work was supported by Research and Development Laboratory, Department of Computer Science and Engineering at Bhilai Institute of Technology, Durg, Chhattisgarh, India, awaiting sponsorship from suitable funding agencies.

6. REFERENCES

[1] Bhardwaj B., Pal S.(2011), "Mining Education data to Analyze Students' Performance", International Journal of Advanced Computer Science and Applications, Vol. 2,No 6, 63-69.

[2] Kumar V., Chadha A.(2012), "Mining Association Rules in Student's Assessment Data" , International Journal of Computer Science issue, Vol. 9,NO. 3, 211-216.

[3] Angeline D.M.D.(2013), "Association Rule Generation for Student Performance Analysis using Apriori Algorithm", The Standard International Journal on Computer Science Engineering & its Applications, Vol. 1, No. 1, 12-16.

[4] Singh M., Singh J., Rawal A.(2014), "Feature Extraction Model to identify At-Risk Level of students in academia", Information Technology (ICIT),2014 international conference on 22-24 Dec 2014, 221-227, ISBN-978-1-4799-8083-3.

[5] Singh M., Singh J.(2013), "Machine Learning Techniques for prediction of subject scores: A Comparative Study", IJCSN, Volume 2,Issue 4, August 2013, ISSN- 2277-5420.

[6] Singh M., Rawal A., Dubey S. (2010),"Modeling C4.5 Decision Tree Classifier for prediction of subject scores in ongoing courses of Academia", NCICT 2010.