Mining the Impact of Course Assignments on Student Performance

Ritu Chaturvedi and C. I. Ezeife*
School of Computer Science, University of Windsor
Windsor, Canada N9B 3P4
rituch@uwindsor.ca and cezeife@uwindsor.ca

ABSTRACT

Educational model for higher education has shown a drift from traditional classroom to technology-driven models that merge classroom teaching with web-based learning management systems (LMS) such as Moodle and CLEW. Every teaching model has a set of supervised (e.g. quizzes) and/or unsupervised (e.g. assignments) instruments that are used to evaluate the effectiveness of learning. The challenge is in preserving student motivation in the unsupervised instruments such as assignments as they are less structured compared to quizzes and tests. The research applies association rule mining to specifically find the impact of unsupervised course work (e.g. assignments) on overall performance (e.g. exam and total marks).

1. INTRODUCTION

Research has shown that society is gradually drifting from the most common teacher-centered classroom teaching model to a hybrid educational model that combines classroom teaching and technology such as internet [2]. Some of the recent technology-based systems are web-based courses, learning content management systems (LMS), adaptive and intelligent web-based educational systems (Intelligent tutoring systems (ITS)) [2] and more recent online systems such MOOC (Massive open online course).Such web-based courses gather student data using activities such as quizzes, exams and assignments to measure their cognitive ability and additional data to measure other factors that could influence learning such as number of times a student has visited a webpage. The objective of this paper is to study the direct and indirect impact that unsupervised tasks (e.g. assignments) have on final marks and grades using association rule mining (ARM). ARM is an unsupervised learning method that looks for hidden patterns in data. An impact on the overall mark is considered to be direct (if the student achieves a score x in the assignment, a 10% of x contributes to the overall mark).An impact on final exam is considered to be indirect – student performing well in the assignments understands the course concepts well and hence performs proportionately well in the final exam. The motivation behind this research is to offer constructive suggestions to educators to help them effectively decide the optimal number of course assignments and the amount of weight that should be given to them (e.g. giving 15% weight to the assignments as opposed to 10%).

2. RELATED WORK

Data mining techniques such as classification, clustering and association rule mining have been used to provide guidance to students and teachers in activities such as predicting student's performance and failure rate, discovering interesting patterns among student attributes and finding students who have a low participation in collaborative work [1,2]. Even though a lot of work exists on mining educational data, no one so far has attempted to analyze the impact that assignments have on student performance in a course. In this paper, we mine student data from courses offered in the University of Windsor, Canada to study this impact.

* This research was supported by the Natural Science and Engineering Research Council (NSERC) of Canada under an operating grant (OGP-0194134) and a University ofWindsor grant.

3. THE PROPOSED SYSTEM FOR MINING STUDENT LEARNING

3.1Learning Management System Overview

CLEW (Collaboration and Learning Environment Windsor) is an LMS developed by University of Windsor and is used to support teaching and learning in face-to-face, distance and blended courses (www.uwindsor.ca/clew). Each course has a set of objectives and a set of supervised (e.g. tests) and unsupervised (e.g. assignments) instruments. Students can use discussion board and chat rooms to collaborate with peers or post concerns.

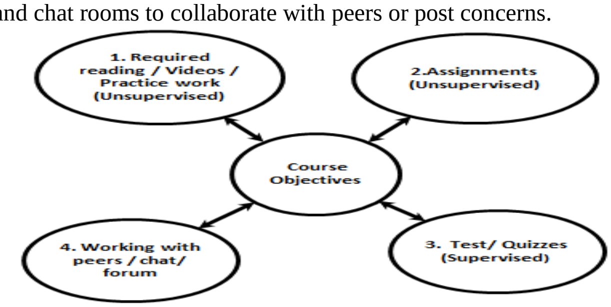

Figure 2: Assessment instruments in courses on CLEW

3.2 Proposed algorithm Mine_Learning (ML):

This section first presents a formal algorithm (Mine_Learning (ML) for mining impact of assignments on student performance before providing more detailed description of each step.

Algorithm Mine_Learning (ML)
Input: an excel file with student marks (individual assignments, average assignment, final, total), number of visits to CLEW
Output: association rules generated for the input dataset, a model predicting if a student can pass a course given assignment and final marks

1. Prepare and clean the student marks.
2. Transform input data into sets of items to get marks_items.
3. Call the Apriori algorithm with marks_items to get marks Frequent patterns and Marks association rules.
4. Interpret the importance of generated mark association rules using confidence and support.
5. Apply SVM to define the Pass/Fail classification model

Figure 3 : Proposed Mine_Learning algorithm

Step 1: Data Cleaning, preparation and preprocessing of marks: Missing marks were replaced by zeroes. A row with 0 as the final exam mark (possibly because the student did not write the final exam) was deleted. Attributes such as ID that were of no significance to this study were removed. All marks were converted to percentages (out of 100) to be consistent.

Step2: Data transformation: Each row of input data is transformed into a set of items as a preparatory step for the Apriori algorithm. For example, let's assume that a student achieves {85, 80, 78} respectively as average assignment, final exam and overall marks. These marks are converted into items {R1, R5, R10} based on the rules defined in table 1. Table 1 defines rules R1 – R4 for assignment marks, R5 – R8 for final exam, R9 – R12 for overall marks and R13 – R15 for number of visits made to CLEW. Rules 16 onwards for individual assignments use the same categories as assignment_average and are not shown here. The transformed data as shown in table 2 has attributes TID (transaction Id), Num_att (number of attributes), RuleAA (Assignment_average transformed as 1 - 4), RuleFM (Final_exam transformed as 5 - 8), RuleTM (Overall_mark transformed as 9 -12) and RuleV (Number of visits as 13 - 15). 'R' has been removed from the transformed items for convenience.

Table 1: **Rules used to transform marks to items**

Rule Identifier	Rule
R1	**Assignment_average> 85**
R2	**70 <Assignment_average<= 85**
R3	**50 <= Assignment_average<= 70**
R4	**Assignment_average< 50**
R5	**Final_exam_mark> 85**
R6	**70 <Final_exam_mark<= 85**
R7	**50 <= Final_exam_mark<= 70**
R8	**Final_exam_mark< 50**
R9	**Overall_mark> 85**
R10	**70 <Overall_mark<= 85**
R11	**50 <= Overall_mark<= 70**
R12	**Overall_mark< 50**
R13	**1 <Number_of_visits<=100**
R14	**100 <Number_of_visits<=200**
R15	**Number_of_visits> 200**

Table 2: Output of step 2

TID	Num_att	RuleAA	RuleFM	RuleTM	RuleV	RuleA1	RuleA2	RuleA3	RuleA4	RuleA5	RuleA6
1	10	2	7	11	16	20	25	31	33	36	40
2	10	1	7	10	16	20	24	30	32	37	40
3	10	4	7	11	16	22	25	30	33	39	43
4	10	2	7	10	16	20	25	31	33	36	41
5	10	3	7	11	17	23	26	30	33	36	43
6	10	4	6	11	16	22	27	31	35	37	43
7	10	3	7	11	16	22	25	31	35	36	41
8	10	2	7	10	16	20	25	31	33	36	40
9	10	2	7	10	16	20	26	29	33	36	41
10	10	1	7	10	17	20	24	28	32	37	40
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

Step3: Apply Apriori algorithm to the dataset obtained as output of step 2 to generate rules such as R4 =>R 8, R4 =>R12, R6 =>R1 and so on, as shown in table 3.

Step4: The rules generated in step 3 are then manually interpreted using the rules' confidence and support values to answer questions such as 'Does average assignment mark have a favorable impact on Final / Overall marks?' or 'How important is it for students to visit the CLEW site frequently?'.

Step 5: An SVM model is created to classify student data (from step 1) as Pass / Fail based on total marks. A total mark of >=50 is labeled as Pass; < 50 is labeled as Fail.

Table 3:Sample rules generated

Rule	Antecedent	Consequent	Support	Confidence
1	Assignment_average < 50	Final < 50	24	63
2	Assignment_average < 50	Total < 50	30	79
3	70 < Final <= 85	Assignment_average>85	16	56

4. RESULTS

Experiments were conducted on student data of two semesters in two Computer Science courses: 'Programming in C for Beginners'(code 106F and 106W) and 'Key concepts for end-users' (code 104F). A relative weight of 30% was assigned to assignments in 106F, 10% in 106W and 50% in 104F.Threshold used for support were 15 and 20 and for confidence was 50 since experiments indicated that lowering minimum support and confidence values increased the number of rules generated substantially. An analysis of all rules generated asserts our hypothesis that assignment marks have an impact on the student's overall and final marks. The confidence of such rules for course 106F is much higher compared to 106W, which implies that allocating a 30% weight to the assignments (as in 106F) has a higher impact than 10 % (106W). Similarly, rules generated for the last 5 assignments in 106F had 100% confidence in depicting that a student who scores > 85 in assignment also scores > 85 in final exam and in the overall mark, and a similar trend was observed with assignment marks <50. Some rare rules such as one found in the dataset 106W (Assignment1 > 85 => Final <50) can be attributed to the fact that assignment1, being the first one, was either too simple or marking was too lenient. Rules generated for all assignments of 104S are uniformly indicative of the fact that achieving a score of > 85 in the assignment and final exam ascertains a total mark in the range of >=70 and < 85. SVM model for all the three courses using average assignment marks predicted a student's chance of passing the course with more than 95% accuracy. However, accuracy using individual assignment marks was 81%, 87% and 97% for 106W, 106F and 104S respectively.

5. CONCLUSIONS

This research presents how useful the association rules mined using Apriori algorithm on student data are in extracting hidden patterns about course assessment instruments such as assignments and final exam. Teachers can take informed decisions using such patterns and use them in improving their curriculum and strategy of teaching a class. The assertion that assignment marks have a direct correlation with final exam and overall marks can be a motivating factor for them to perform well in the assignments.

6. FUTURE WORK

Moving forward, we propose to mine more student attributes that could impact their learning such as their personalities and meta-cognitive skills. Other instruments such as social aspects using chat rooms, discussions and forums that are prevalent in today's web-based courses can also be mined to study their impact on learning.

7. REFERENCES

[1] Anwar, M., and Ahmed, N. Knowledge mining in supervised and unsupervised assessment of students performance. In 2nd International Conference on Networking and Information Technology, 2011, 29-36.

[2] Minaei-Bidgoli, B. Data mining for a web-based educational system. PhD thesis, Michigan State University, 2004.