

Accepting or Rejecting Students' Self-grading in their Final Marks by using Data Mining

J. Fuentes

Department of Computer Science, University of Cordoba, Spain

javier.fuentes@uco.es

C. Romero

Department of Computer Science, University of Cordoba, Spain

cromero@uco.es

C. García-Martínez

Department of Computer Science, University of Cordoba, Spain

cgarcia@uco.es

S. Ventura

Department of Computer Science, University of Cordoba, Spain

sventura@uco.es

ABSTRACT

In this paper we propose a methodology based on data mining and self-evaluation in order to predict whether an instructor will or will not accept the students' proposed marks in a course. This is an on-going work in which we have evaluated the usage of classification techniques and cost-sensitive corrections. We have carried out several experiments using data gathered from 53 computer science university students.

Keywords

self-grading, self-evaluation, cost-sensitive classification.

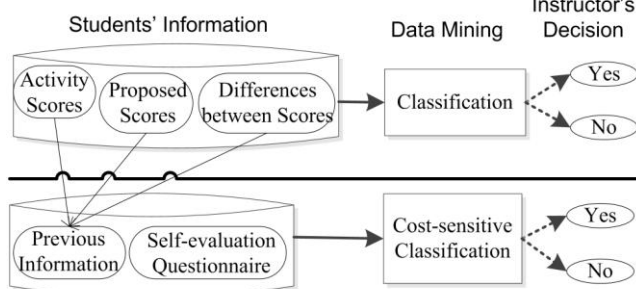
1. INTRODUCTION

Assigning appropriate grades to students is an arduous and difficult process for instructors. Grades are, by their nature, somewhat subjective; every instructor uses different criteria to assign them and place a different emphasis on them. And with trends in higher education moving toward large class sizes, yet simultaneously toward more personalised and individualised instruction, self-grading may facilitate the achievement of these two objectives [3]. However, the main disadvantage of self-grading is grade inflation, that is, normally, more students, particularly among younger students, grading themselves higher than what they should get [4]. Roughly speaking, students' self-gradings are satisfactory substitutes for teacher gradings, if these two measures are comparable. If a student's grades were very different from the teacher's judgment, then the teacher should supervise and thoroughly evaluate the work, activities, and/or exams. Following this idea, in this paper we are interested in predicting what the instructor's decision is concerning the possible acceptance of the students' proposed final marks in a course. To do that, we use a methodology based on classification and self-evaluation checklists [1].

2. METHODOLOGY

The methodology that we have used in this study is as follows. During the course, students are evaluated by means of a multiple choice testing that is an effective assessment technique. Before the final exam date, all students are requested to self-grade. Students propose the mark/grade that they think they should get for the course. Then, the instructor accepts or declines the proposed mark of each student as the final mark for the course. This way, only students whose score was declined by the instructor will have to sit the final exam. Finally, we try to predict the instructor's decision of accepting or declining the score proposed by students. We have used two different (initial and new) data mining approaches (see Figure 1).

INITIAL APPROACH



NEW APPROACH

Figure 1. Approaches for predicting instructor's decision.

The initial approach uses three numerical variables: the score obtained by students in the course's activities, the proposed scores by students and the difference between these two previous scores. Then, it applies traditional classification algorithms for predicting the instructor's decision about whether to accept the proposed students' scores (YES) or not (NO). The new approach uses the three previous variables as well as a self-evaluation questionnaire as another source of information. Then, it applies cost-sensitive classification [2] that is normally used for obtaining better performances than traditional classification with unbalanced datasets. In fact, in our particular problem we are much more interested in the correct classification of NO (normally the minority class) than YES (the majority class). To do that, costs can be incorporated into the algorithm and considered during classification. In the case of two classes, costs can be put into a 2×2 matrix in which diagonal elements represent the two types of correct classifications and the off-diagonal elements represent the two types of errors. This matrix indicates that it is N times more important to correctly classify NO than YES students.

3. DATASET

We have used a dataset collected from second year university Computer Science students in 2012-13. During a traditional, face-to-face course on artificial intelligence, the instructor gave the students the option to self-grade. Out of the 86 students enrolled in the course, 53 accepted to self-grade, approximately 60%. For each one of these 53 students, we gathered the next attributes:

- **Activities score.** This is the average score obtained by students in three activities undertaken during the course. The three activities were Moodle multiple-choice tests with 10 questions, available at different moments of the course. The activities score of each student is a number between 0 and 10 points that is the average of the three activities.

- **Proposed score.** This is the final mark/score that the students believe that they should get in the course. Students themselves proposed their marks (number between 0 and 10).
- **Difference between scores.** It is the difference between the two previous scores. It is a positive or negative value (between -10 and +10) obtained automatically as the activities score minus the proposed score.
- **Self-evaluation questionnaire score.** This is the score obtained in a self-evaluation questionnaire. We have used a self-evaluation questionnaire developed at the University of Ohio (USA) [5]. It contains 50 yes/no questions for determining whether a student is a good or poor student. The students completed the questionnaire two weeks before the final exam date. The University of Ohio also provides a template with the responses of good students. Using this template we have calculated a score for each student as the number of answers equal to those of the good students.

The output attribute or class to predict in our problem is the instructor's decision. It is a binary value: YES or NO, that indicates whether the instructor accepts or declines the students' proposed scores. The instructor provided us with this value for each one of the 53 students: 37 YES (70%) and 16 NO (30%).

4. EXPERIMENTS

We have carried out several experiments in order to test our proposed methodology for predicting the instructor's decision. In these experiments we have used 35 classification algorithms provided by Weka 3.7: NaiveBayes, NaiveBayesSimple and NaiveBayesUpdateable, Logistic, RBFNetwork, SimpleLogistic, SMO, SPegasos, VotedPerceptron, MultilayerPerceptron, IB1, IBk, KStar, LWL, ConjunctiveRule, DecisionTable, DTNB, JRip, NNge, OneR, PART, Ridor, ZeroR, ADTree, BFTree, DecisionStump, FT, J48, J48graft, LADTree, LMT, NBTree, RandomForest, RandomTree, REPTree and SimpleCart. We executed all the algorithms using 10-fold cross-validation and their default parameters. Three classification performance measures were used to test the algorithms' results: Accuracy, True Positive rate (TP rate) or sensitivity, and True Negative rate (TN rate) or specificity. Figure 2 shows the obtained average values of all algorithms when using the initial and new approach with different values of cost ($N = 1, 2, 3, 4$ and 5).

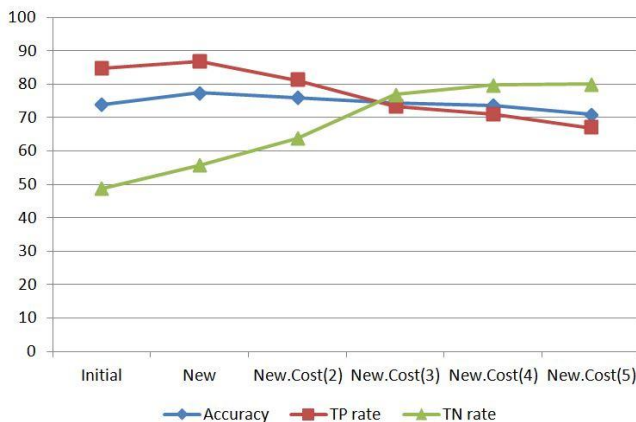


Figure 2. Traditional/Initial classification versus New/Cost-sensitive classification performance.

We can see in Figure 2 that the new approach improved the initial approach in the three evaluation measures and so, the self-evaluation questionnaire has shown to be a good source of information. However, TN rate continued at a very low values and so, we applied different costs for improving it. In Fact, Figure 2 also shows that when we increase the cost/weight of correctly classified NO students, it increases the TN rate. However, it also decreases the accuracy and TP rate. So, it is necessary to select the best N value in our problem in which TN rate improves without affecting accuracy and TP rate very much. For example, in our case, we can see that in Figure 2 an agreement/good solution is for $N=3$ in which the three measures cross its values.

Finally, we show an example of a model obtained with one of the classification algorithms. We have selected the output of the J48 algorithm (see Figure 3) because it obtained one of the best performances and it is also a well-known white-box classification algorithm. Using the discovered IF-THEN rules, instructor can make decision about which students accept or not their scores.

```

IF Difference-Between-Scores >= 0 THEN Decision=YES
ELSE IF Difference-Between-Scores < 0
  AND Proposed-Score <= 5 THEN Decision=YES
  ELSE IF Proposed-Score > 5
    AND Self-evaluation-Score >=5.9 THEN Decision=YES
    ELSE IF Self-evaluation-Score <5.9 THEN Decision=NO

```

Figure 3. Example of obtained decision tree.

5. CONCLUSIONS

Regarding the performance of the prediction of the instructor's decision, the results obtained show that the use of self-evaluation, and cost-sensitive classification improved the accuracy, sensitivity and specificity in our dataset. Our final objective is to use it as a time-saving scheme because only the rejected students (whose score the instructor does not accept) have to sit the final examination at the end of the course. Currently, we are carrying out more experiments with a greater number of students of different university courses. In the future, we want also work in the calibration task, which refers to how accurately individuals can predict how well they do on a task.

6. ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science Technology project, TIN-2011-22408, and FEDER funds.

7. REFERENCES

- [1] Andrade, M.R. Monitoring student performance with self-evaluation checklists: an ongoing case study. <http://www.jrc.sophia.ac.jp/courses/pdf/kiyou2701.pdf>
- [2] Elkan, C. The foundations of cost-sensitive learning; International Joint Conf. on Artificial Intelligence, 1-6. 2001.
- [3] Strong, B., Davis, M., Hawks, V. Self-grading in large general education classes. *Colleague Teaching*, 52(2), 52-57. 2004.
- [4] Tan, K. Qualitatively different way of experiencing students self-assessment. *Higher Education Research&Development*. 52: 52-57. 2008.
- [5] Test de Autoevaluación para Estudiantes Secundarios y Universitarios. <http://www.tecnicas-de-estudio.org/general/sabes-estudiar.htm>