Comparing classification methods for predicting distance students' performance

Workshop on Applications of Pattern Analysis 2011

Diego Garcia-Saiz

Marta Zorrilla

Contents

- Introduction to Educational DataMining (EDM)
- Research in EDM at University of Cantabria: previous work
- Comparing classification methods for predicting distance students' performance
  - Objetives
  - Related Work
  - Data and Algorithms
  - Results
  - A new approach: Pre-processing data with classification algorithms
  - Conclusions and Future Work

Introduction to Educational DataMining (EDM)

- **“Educational Data Mining** is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.” (Educational DataMining Society).

- It exists an International Group, a Journal and an annual Conference in EDM.

Research in EDM at University of Cantabria: previous work

- E-learning WebMiner:
  - A web application that offers the opportunity of getting data mining models to instructors involved in e-learning courses .
    - What tools are used together.?
    - What kind of sessions are carried out in the course?.
    - What are the students' profiles?.
- Two posters in EDM 2011:
  - E-learning WebMiner
  - Yacaree in EDM

Comparing classifiers to predict the success of e-learning students

- Objetives:
  - Compare different classifiers to determinate which is the best algorithm to predict the success of e-learning students.
  - Establish if the choosen algorithms are easy to interpret by non-expert users.
  - Develop a new preprocesing algorithm that improves the results obtained in the previous comparison
  - Integrate the classification algorithms and the preprocesing algorithm in ElWM to offer new templates.

Related work

- There are a few previous works and papers written by other authors that discuss and compare differente classificacion algorithms in educational datamining context.

  - Common feature: predict students’ performance

  - Differences between these reseach and ours:

    - Consider all kind of courses, not only e-learning
    - Don’t focus in non-expert users or instructors that don’t have knowledge in datamining

“Introduction to Multimedia” course

- We used the data from a course entitled "Introduction to multimedia methods“:
  - Open to all degrees.
  - The average number of students enrolled per year is about 70.
  - Only 25% of students deliver all the tasks…
    - … a lower number pass the course.
  - As consequence of that fact, the class attribute is unbalanced.

Datasets and attributes

- Three different datasets:

Name	Nº of instances	Nº of attributes	Kind of attributes
Mu0710	194	6	Numeric
Mu0910	65	6	Numeric
Mu0910S	65	11	Numeric and categorical

- Attributes:

Numeric	Categorical
Total time spent	Active/Refletive
Numer of sessions carried out	Sensitive/Intuitive
Number of sesions per week	Visual/Verbal
Average time per week	Sequential/Global
Average time per session	Degree

- The class attribute is the Mark (pass or fail the course)

Algortihms

- Rule out the use of Artificial Neural Networks and Support Vector Machine

- The chosen algorithms were:
  - Rule-Based algorithms: OneR
  - Decision Trees: J48
  - Bayesian: Naïve Bayes and BayesNet TAN
  - Instance-based learner classifiers: NNge

- The tool Weka was used to perform the analysis

Results: Mu0910

- We set the algorithms with deafult parameters.
  - 10-fold cross validation for estimating generalization performance
  - T-test for comparing the models generated using a significance level of 10%

Table 1: Accuracy and rates of Mu0910 dataset

Algorithm	TPRate	FPRate	TNRate	FNRate	Accuracy
OneR B 3	0.66	0.47	0.53	0.34	65.79
J48 C 0.25	0.74 v	0.17 *	0.83 v	0.26 *	74.21 v
Naïve Bayes	0.77 v	0.25	0.75	0.23 *	77.29 v
BayesNet TAN	0.76 v	0.18 *	0.82 v	0.24 *	76.36 v
NNge	0.70	0.41	0.59	0.30	70.10

v,* statistically significant improvement or degradation with respect to OneR

Results: Mu0710

- BayesNet performs better than the rest, followed very closed by J48

Table 2: Accuracy and rates of Mu0710 dataset

Algorithm	TPRate	FPRate	TNRate	FNRate	Accuracy
OneR B 6	0.78	0.20	0.80	0.22	77.86
J48 C 0.25	0.79	0.17	0.83	0.21	79.36
Naïve Bayes	0.76	0.27 v	0.73 *	0.24	76.40
BayesNet TAN	0.81 v	0.15 *	0.85 v	0.19 *	81.26 v
NNge	0.78	0.22	0.78	0.22	78.04

v,* statistically significant improvement or degradation with respect to OneR

Results: Mu0910S

- NaïveBayes performs better than the others, and BayesNet is one of the worse algorithms.

Table 3: Accuracy and rates of Mu0910S dataset

Algorithm	TPRate	FPRate	TNRate	FNRate	Accuracy
OneR B 3	0.65	0.48	0.52	0.35	65.29
J48 C 0.25	0.76 v	0.23 *	0.77 v	0.24 *	75.83 v
Naïve Bayes	0.81 v	0.20 *	0.80 v	0.19 *	80.90 v
BayesNet TAN	0.73	0.29	0.71	0.27	73.36
NNge	0.75	0.36	0.64	0.25	74.98

v,* statistically significant improvement or degradation with respect to OneR

New Approach: pre-procesing with classifiers

- There are outliers in datasets:
  - Students with one learning session can pass the course and students with a high time spent in the course fail.

- We build a meta-algorithm to eliminate these outliers.
  - This consists of building an initial classifier using 10-fold cross-validation and eliminating, randomly, a certain percentage of the instances which belong to the worse classified class, and next, building a second classifier with the filtered dataset using also 10-fold cross-validation.

Results with the pre-processing algorithm

- Accuracy is higher when we use this meta-algorithm:
  - Applying Naïve Bayes on Mu0710 dataset in both stages, the pass and fail rates increased from 63.75% to 75.00% and from 85.09% to 86.84% respectively.
  - Applying J48 in both stages, the pass and fail rates increased from 91.25% to 95.00% and from 71.93% to 86.81% respectively

Results with the pre-processing algorithm: another approach

- Better results were got when the instances ruled out were chosen taking into account the value of the most significant attribute for the algorithm.

  - Applying Naïve Bayes in both stages, got an improvement in the fail rate from 85.09% to 87.72% and got worse in the pass rate, from 75.00% to 71.88% in comparison with the previous results.

  - But, applying J48 in both stages, both rates increased, the pass rate from 91.25% to 100% and the fail rate from 71.93% to 86.81%.

Results with the pre-processing algorithm: another approach (more)

- With J48 we eliminated the bad classified "fail" instances, whereas, with Naïve Bayes, we ruled out the bad classified "pass" instances.
  - When we eliminated the bad classified "fail" instances with Naïve Bayes we achieved an improvement in pass and fail rates from 63.75% to 90.72% and from 85.09% to 91.25% respectively.
    - We obtained better results than in the previous tests carried out with Naïve Bayes.
- Another alternative tested consisted on applying this preprocessing in each cross-validation fold of the first classifier…
  - … But, in this case, the instances ruled out in each fold were of different class and the results obtained were worse than the previous

Conclusion and Future Work

- Our experimentation shows that there is not one algorithm that obtains a significantly better classification accuracy.
  - NaïveBayes performs better with smaller datasets or datasets which have both categorical and numeric attributes with missing data.
  - J48 performs better with bigger datasets, close to BayesNet…
    - But BayesNet models are not easy to interpret.
- Both Naïve Bayes and J48 have better results if the preprocessed task is carried out according the most significant attribute for the algorithm used to preprocess.
- Our near future work is to extend this experimentation with other datasets to validate these conclusions and next, to study how the meta-algorithm can set itself according to the dataset and, lastly, to add a template for predicting the students' success in our tool.