Performance Indicators for Online Geography in Secondary Education

van Diepen, P.; Bredeweg, B.

Publication date
2016

Document Version
Final published version

Published in
BNAIC 2016 : Benelux Conference on Artificial Intelligence

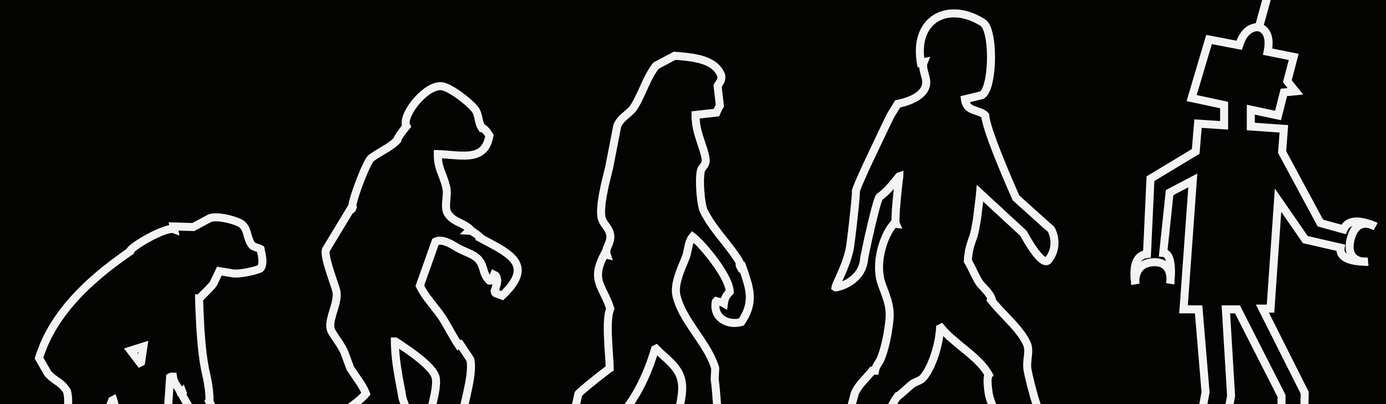
Citation for published version (APA):
van Diepen, P., & Bredeweg, B. (2016). Performance Indicators for Online Geography in Secondary Education. In T. Bosse, & B. Bredeweg (Eds.), *BNAIC 2016 : Benelux Conference on Artificial Intelligence: proceedings of the Twenty-Eight Benelux Conference on Artificial Intelligence : Amsterdam, November 10-11, 2016* (pp. 206-207). (BNAIC; Vol. 28). Vrije Universiteit, Department of Computer Sciences.
http://bnaic2016.cs.vu.nl/index.php/proceedings

BNAIC2016

PROCEEDINGS

OF THE 28TH BENELUX CONFERENCE ON ARTIFICIAL INTELLIGENCE

NOVEMBER 10-11 2016
AMSTERDAM



Performance Indicators for Online Geography in Secondary Education[1]

Pepijn van Diepen
pepijnvandiepen@gmail.com

Bert Bredeweg
B.Bredeweg@uva.nl

Informatics Institute, Faculty of Science
University of Amsterdam, The Netherlands

Abstract

There is little consensus about what variables extracted from learner data are the most reliable indicators of learning performance. The aim of this study was to determine those indicators by taking a wide range of variables into consideration concerning overall learning activity and content processing. A Genetic Algorithm was used for the selection process and the variables were evaluated based on their predictive power. Variables extracted from exercise activities turned out to be the most informative.

1 Introduction

The analysis of learner data can provide insight into the progress of students and their learning performance. Students and teachers can benefit from feedback on the learning process whereas publishers could benefit from feedback on their content [1]. The prediction of learning performance can support this, however, scientists are not confident about what data is most suited for this purpose [1, 6]. Quantitative data concerning *resource use*, *time spent on resources* and *grades* have been used for the prediction of learning performance in the reviewed literature [4, 6]. The aim of this study was to determine what aspects of learning behavior can be extracted from a Learning Management System (LMS) in secondary education and are reliable indicators of learning performance.

2 Method

The data for this research was provided by educational publisher ThiemeMeulenhoff and was extracted from the online geography course De Geo[2]. It consisted of chronological activity logs and exercise results of 226 first grade secondary education students. The course material included reading material (also referred to as *theory*), online exercises and self-assessment tests. Each exercise was categorized according to Bloom's taxonomy for learning objectives [3]. Exercise activity was therefore analyzed separately for each category. Since all data was anonymous and no final grades were made available due to privacy constraints, learning performance had to be determined based on alternative sources. The self-assessment tests were designed to provide the students an indication of their learning performance, therefore results on self-assessment tests were found to be the most appropriate measure of learning performance. First, variables concerning overall online activity were considered (e.g. number of clicks, time online, theory/exercise time spent ratio). Subsequently, content specific variables extracted from reading and exercise activities were analyzed. All data was categorized in terms of exercise processing and theory processing. Subsequently an extensive set of variables was composed for each category

[1]The resources for this project were provided by ThiemeMeulenhoff. Special thanks go to Joost Borsboom, Gilian Halewijn, Wouter van Rennes, Emiel Ubink and Johan Verhaar (in alphabetic order). The full BSc thesis can be found at: `https://staff.science.uva.nl/b.bredeweg/pdf/BSc/20152016/Van%20diepen.pdf`

[2]De Geo is part of Edition by ThiemeMeulenhoff. More information: `https://www.thiememeulenhoff.nl/voortgezet-onderwijs/mens-en-maatschappij/aardrijkskunde/de-geo-onderbouw-9e-editie`

based on the type of variables that were found to be reliable in the reviewed literature. Reading activity variables concerned the number of clicks and time spent. Exercise activity variables covered the number of incompletes (wrong answer), completes and the time spent. A selection was made using a univariate variable selection method based on linear correlation with learning performance. Subsequently, multivariate variable selection was applied on the remaining variables using a Genetic Algorithm (GA). By using the prediction performance as fitness and variable subsets as individuals, the GA selected the strongest combination of variables. Since usage of GAs for variable selection is prone to overfit [2], the entire selection process was 10-fold cross validated. Each fold had an optimal variable subset as output and a voting mechanism was used to make a final selection. The predictive power of the selected variables was evaluated in two learning performance classification tasks: fail/pass and fail/sufficient/excellent. Classification algorithms such as Support Vector Machine (SVM), Gaussian Naive Bayes (GNB) and K-Nearest Neighbour (KNN) provided by the Scikit-Learn library [5] were used. All classifications were evaluated using repeated 10-fold cross validation. The accuracy of the classifications was examined alongside with the recall of the poor performing class. Finally, a baseline classifier that was set to always predict the most common class was included in the evaluation.

3 Results and Conclusion

The selection process resulted in a final set of fifteen variables coming from almost all categories. Two belonged to the overall activity category: the total number of clicks and the exercise-theory time spent ratio. Two variables belonged to the reading activity category: the number of theory accesses and amount of time spent on theory while doing exercises. Eleven variables were of the exercise activity category, especially exercises from the *apply* category (five variables) and *understand* category (three variables) from Bloom's taxonomy were found to be reliable indicators. No variables came from the *remember* and *analyze* category and one variable from both the *evaluate* and *create* category. From the exercise processing variables the number of incompletes (wrong answer provided) was most informative, followed by the mean and total time spent on exercises. The last variable of the list concerned the mean time spent on an exercise over all of Bloom's categories together.

The baseline classifiers achieved an accuracy of 0.51 and 0.50 in the two and three class classification task respectively. The SVM classifier predicted most accurately for both classification tasks followed by GNB and KNN. An accuracy of 0.80 and recall of 0.84 of the *fail* class was achieved at the classification of two classes. For the classification of three classes an accuracy of 0.67 and recall of 0.67 was achieved. Other classification algorithms were also evaluated but resulted in less accurate predictions. However, all classifiers did perform significantly better (0.05) than the baseline classifiers.

Concluding, the indicators concerning exercise processing were found to be the most reliable. Variables extracted from exercise activities that were designed to train students in *understanding* and *applying* material were found to be especially informative. However, a combination of features concerning overall activity, theory- and exercise-processing was needed to achieve the best prediction results. Therefore it is important to capture as many aspects of the learning process as possible in order to make accurate predictions.

References

[1] J. Esmeijer and A. van der Plas. Learning analytics en zelfsturend leren. *TNO R10373*, 2013.

[2] I. Guyon et al. An introduction to variable and feature selection. *JMLR*, 3:1157–1182, 2003.

[3] D.R. Krathwohl. A revision of bloom's taxonomy: An overview. *Theory Pract*, 41(4):212–218, 2002.

[4] L.P. Macfadyen and S. Dawson. Mining lms data to develop an early warning system for educators: A proof of concept. *Comput Educ*, 54(2):588–599, 2010.

[5] F. Pedregosa et al. Scikit-learn: Machine learning in Python. *JMLR*, 12:2825–2830, 2011.

[6] D.T. Tempelaar, B. Rienties, and B. Giesbers. In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Comput Human Behav*, 47:157–167, 2015.