Mining students' data to identify at-risk students in an academic English course: A comparison of two classification techniques from a language teacher and statistical perspective

Dennis FOUNG (dennis.foung@gmail.com)
The Hong Kong Polytechnic University, Hong Kong

Abstract

In this study, mixed methods were used to explore the effectiveness of data-mining techniques from statistical and language teacher perspectives. This study is important, because comparison of data-mining techniques has seldom been conducted in the higher education language-learning context. In addition, not many previous comparison studies considered the perspective of language teachers, as the ultimate user of the results. This study used a data set with more than 5,000 students from two academic courses offered at a university in Hong Kong, and adopted two commonly used data-mining techniques: classification tree and logistics regression analysis. This quantitative analysis explored the suitability of data-mining techniques. To understand the language teacher perspective of these techniques, the results were presented to a group of 16 professional English teachers, to check whether they thought the results were useful. Results showed that despite satisfactory results in both data-mining techniques, the teachers were very hesitant to use them. The teachers' resistance stemmed from their doubts about the techniques, and the applicability of these techniques in the language education context. Further research should be conducted to promote these techniques to language teachers.

Keywords: data mining, classification techniques, at-risk students, early warning system

Introduction

Language learning is seldom considered an independent process, as it requires feedback and guidance from teachers or computers. Students may not know how well or badly they are performing in their assessments. This feedback process is important at the microlevel, in a classroom (or in front of a computer-assisted language-learning program), to let students know how they are performing in an activity, and at the macrolevel, in courses, to let students know how well they are progressing. In practice, such feedback may be difficult to facilitate. For example, in higher education, students may attend language classes with many students (Abdullah, Ramli, & Refek, 2017), but they still need teacher feedback about their progress. However, individualized feedback from teachers may not be possible in larger classes (Chingos & Whitehurst, 2011). To facilitate this feedback process, scholars (and even companies) have attempted to develop different feedback systems to identify weaker students and provide actionable feedback (i.e., feedback items that students can take proper action with) to these weaker students. These systems make use of data-mining techniques (enabled by computers) to predict students' performance in courses (Park, Yu, & Jo, 2016). This motivates the present study to explore which data-mining techniques are preferred from statistical and language teacher perspectives.

Context

The research site was a university in Hong Kong that admits mainly local students. These students have completed secondary school, and most have achieved Level 3 or 4 in the secondary school exit exam (see the Methods section for more details). The city (and the university) advocates the use of smart systems and smart technologies, and students and teachers are no strangers to the latest technological terms, such as data-mining and analytics. However, other than the functions in the learning management system, this research site is not adopting any university-wide smart systems for teaching at the research site (e.g., a dashboard).

Significance

This study makes an important contribution to language-learning research, because this research on data-mining techniques is uncommon, and much data-mining research has not included language teacher perspectives. Although some research in this field used advanced statistics, many language-learning scholars used traditional hypothesis-testing methodologies, such as Alfehaid (2018) and Kongsuebchart and Suppasetseree (2018). These studies made important contributions to the field. However, with the emergence of advanced technologies, large data sets are available, and data-mining techniques provide an alternative to traditional statistics research. It would be helpful if the usefulness of data-mining techniques were explored. Furthermore, studies on predictive analytics are common in educational research (such as Bainbridge et al., 2015); however, they often focus on the accuracy of these techniques, and the language teachers (as potential users) perspective has often been neglected. If the potential users' perspective is neglected, the actual impact of a statistically useful data-mining technique cannot be properly estimated. This shows the importance of this study.

Literature Review

At-Risk Students

The main reason for using data-mining techniques, and for conducting this research, was to identify at-risk students. Unlike high schools with a limited number of students in classes, mass lectures are widely adopted in college classrooms, including language classrooms (Abdullah et al., 2017). In classrooms with many students, teachers may not have an opportunity to interact with most students during lessons. In particular, teachers may not know the students' progress and give student-specific feedback to students (Chingos & Whitehurst, 2011). It is possible that students may not be progressing well; and have the possibility of failing the course. Therefore, these students are at-risk students.

Previous studies in other contexts defined at-risk students as those who may fail a course (such as Conijn, Snijders, Kleingeld, & Matzat, 2017). In the tertiary language learning context in Hong Kong, the student dropout rate is very low (Kwok, 2016), and there is no final exam. Therefore, in this study, at-risk students were defined as those who could not achieve a good grade, which is similar to Jayaprakash, Moody, Lauría, Regan, and Baron's (2014) definition of at-risk students. As, in this study, at-risk students simply refer to those who are weaker in performance, "at-risk" students and "weaker" students are used interchangeably in this article.

Data Mining

Data mining has been popular for years (Sahoo, 2013). Techniques such as the neural network (the backbone of deep learning) were proposed as early as the 1940s, but the computers at that time were not able to handle such computation-intensive processes. With advanced technology, computers are now able to handle massive data and their computation processes, and this has given rise to computation-intensive data mining. Therefore, data mining and related studies fall within the spectrum of computer-assisted research.

In data mining, mining is described as a "knowledge discovery process" (Al-Maqaleh & Abdullah, 2017; Finlay, Pears, & Connor, 2014). In other words, the notion of discovery implies that data mining is actually an attempt to discover, rather than prove, something. In practice, researchers retrieve large but readily available data sets from databases directly. With a general direction (and no specific purpose), researchers attempt to look for patterns and relations from various variables in a data set. This discovery process is quite different from traditional quantitative research, which is led by hypothesis-driven statistics (Baepler & Murdoch, 2010). In traditional research, a specific research hypothesis is formulated, and data are then collected based on that hypothesis. This data collection is followed by rigid data-cleaning and -checking processes that eventually lead to hypothesis tests. These tests can tell whether a hypothesis is supported or refuted. Due to the availability of data in the era of big data, data mining has become less costly when compared to the steps necessary for traditional quantitative research, and this technique has been widely adopted in different contexts. Generally, data mining can achieve a number of purposes, as suggested by Goyal and Vohra (2012). See also Romero and Ventura (2010), for a comprehensive review of data-mining techniques in an educational context. These purposes are summarized in Table 1.

Table 1

Purposes of Data Mining

Data-mining techniques and purposes	**Definitions by Goyal and Vohra (2012)**	**Possible applications in language learning**
Anomaly detection (outlier/change/deviation detection)	Identification of unusual data records that might be interesting or of data errors that require further investigation	Identify students who unexpectedly perform too well or too badly on the final assessment
Association rule learning (dependency modeling)	Method for finding relations between variables	Examine relations between course materials and final grades of students
Clustering	Discovery of groups and structures in the data that are similar in some way without using known structures in the data	Group the usage patterns of students in a particular CALL program
Classification	Generalization of known structures to apply to new data	Predict student performance
Regression	Technique that attempts to find a function that models the data with the least amount of error	Establish relations between the use of different online materials and the students' final grades
Summarization	A compact representation of the data set, including	Visualize students' usage patterns in the learning management system in a

	visualization and report generation	course (e.g., which day is the most popular day of access)

Classification Techniques

Classification techniques are very popular. Goyal and Vohra (2012) defined them as "a generalization of known structures to apply to new data." In other words, classification techniques can identify relations between class membership and variables based on the existing data set, to predict the membership of variables in another data set. In practice, researchers can use classification techniques to explore the relation between award classification (e.g., first-class honor or second-class honor as class memberships) and the number of hours spent on the learning management system in the first year. Researchers can explore such relations from a data set with graduates in 2018, and understand how the number of hours in the first year can predict students' final award classifications. Next, researchers can use this relation to predict whether the current first-year students (i.e., first-year students in 2018) can get first-class honors 4 years later. As a machine-learning task, Tan, Steinbach, and Kumar (2014) defined classification as a task of learning a target function (i.e., equation) that can help predict class membership with a given set of attributes. Classification techniques is a category for techniques serving the same function, and many data-mining techniques fall into this category. Dreiseitl and Ohno-Machado (2002) summarized a number of them (Table 2).

Table 2
Classification Techniques

Classification techniques	Dreiseitl and Ohno-Machado (2002)
Support vector machine	Model that builds optimal boundaries between data sets by solving a constrained quadratic optimization problem
k-Nearest neighbor	Algorithm that uses the data directly for classification without building a model first
Decision tree	Algorithm that repeatedly splits the data set according to a criterion that maximizes the separation of the data, resulting in a tree-like structure
Logistic regression	The drawing of a regression line to predict the probability of success of a binary variable
Artificial neural network	System that attempts to model the capabilities of the human brain

Comparison of Previous Studies on Data-Mining Methods

Previous studies have examined the use and effectiveness of these techniques in different contexts, but most of these studies were review articles. For those conducted with empirical data, even fewer were completed within the higher education or language-learning contexts. For example, Paliwal and Kumar (2009) and Shahiri, Husain, and Rashida (2015) conducted only a review to compare the effectiveness of different classification techniques. Dreiseitl and Ohno-Machado (2002) compared the use of logistics regression and artificial neural networks with a biomedical data set. They believed that each technique has its strengths and weaknesses, and that it is difficult to draw a conclusion. Sarle (1994) reminded researchers that the techniques do not compete with each other. However, Dreiseitl and Ohno-Machado (2002) managed to further comment about different classification techniques: The quality of

prediction often depends on the quality of the data set and how the variables are being manipulated. This seems to once again suggest the need to compare the effectiveness of data-mining techniques with an authentic language-learning context. Furthermore, most papers focused on the technique, without considering the perception of language teachers, the actual users, suggesting that a user perspective should be included when comparing these techniques. In particular, this paper aimed to answer the following questions: (a) Which data-mining techniques can make a more accurate prediction of weaker student performances in the language-learning context? (b) Do language teachers have any preferences for the use of data-mining techniques?

Method

Overview

This study examined the application of data-mining techniques to identify at-risk students from statistical and teacher perspectives. In Phase 1 of the study, two commonly used data-mining techniques were used with a data set with students from an undergraduate academic literacy program. The accuracy of these techniques was examined. In Phase 2 of the study, we investigated the attitudes of teachers toward these data-mining techniques. Teacher questionnaires were analyzed.

Participants

Phase 1 (data mining). Participants in Phase 1 of the study were 5,993 undergraduates at a university in Hong Kong. They completed the English academic literacy course offered by the research site. These students had completed the secondary school exit exam, and had achieved a score of around IELTS (International English Language Testing System) 6.31 to 6.51. They entered the research site for an undergraduate program. The students were required to enroll in two English courses for their program: one basic academic literacy course (BALC) and one advanced academic literacy course (AALC). This study attempted to identify the weak students in the AALC (based on the final grades for the AALC) using the assessment results from the BALC and the first assessment of the AALC.

Phase 2 (questionnaire). Participants in Phase 2 were 16 professional English teachers at the research site. Although no demographic information was collected, some general descriptions of staff members at the research site can be offered. Generally, there was a mix of staff profiles in terms of gender, years of experience, and level of interest in data mining. Some were local Hong Kong teachers, but some were native English teachers from different locations throughout the world. All teachers had a basic understanding of the BALC and the AALC, and most had experience teaching at least the BALC or the AALC. They were recruited because they attended an annual event at the research site and were willing to complete the questionnaire voluntarily.

Courses and Assessments

The BALC and the AALC were both offered to students at the research site. The BALC aims at providing basic academic literacy training, from essay writing and referencing skills to academic presentation, while the AALC aims at offering advanced academic literacy training, such as advanced research skills, critical reading and oral defense.

The BALC has three assessments: an in-class problem-solution essay, a discursive essay assignment, and an academic presentation. The AALC also has three assessments: a draft position argumentative essay, an oral defense, and a final position argumentative essay. Each assessment was evaluated based on four components. Components for writing were content, organization, language, and referencing, and those for speaking were content, delivery, language, and pronunciation. Students received grades for each component to derive final grades according to a common university assessment scheme. Table 3 presents the details.

All students (n=5993) took a comparable BALC and AALC. The assessment tasks (types of essays / presentations), requirements (e.g. word length, number of in-text citations), assessment components (See Table 4), and assessment rubrics are exactly the same among these students. The university also has a rigid quality assurance process for assessments, including assessment standardization, moderation, and double-marking procedures; and the marking standard is comparable across these students even though the assessments were marked by different teachers. However, it should be noted that students in these courses can use different referencing styles (e.g. APA, Harvard, IEEE, and Vancouver) in their essays. Also, there are minor changes to the content of the online packages but there was no change to the requirement and format to these online tasks. Still, these should not affect the validity and reliability of the data mining procedures.

Table 3
Assessment Scheme

Grade	Meaning	Corresponding scale
A+	Outstanding	4.5
A		4.0
B+	Good	3.5
B		3.0
C+	Satisfactory	2.5
C		2.0
D+	Adequate	1.5
D		1.0
F	Fail	0,0

Phase 1: Data Retrieval, Cleaning, and Analysis

The data set for data mining went through several processing and cleaning steps before analysis. The support staff at the research site helped retrieve the student assessment data from the learning management system, and assessment results of the students across the two courses were merged. Next, students who did not complete the two courses (e.g., those who were missing one assessment or more) were removed from the data set. The final grades were also converted into binary variables, with B grades or below in one group and the remaining grades (i.e., B+, A, A+) in another group. A total of 5,993 students were in the finalized data set. As a standard procedure for data mining (suggested in Llorente & Morant, 2011), the data set was then divided into a training data set (n = 4,195) and a testing data set (n = 1,798). Next, two rounds of data-mining procedures were conducted with the data set.

Variables for Analysis

There were a total of 18 independent variables and one target variable. Table 4 presents a list of variables analyzed in this study.

Table 4
Assessment Components/Research Variables

Course assessments and assessment components	**Variable type (range)**	**Acronym**
First English Course (BALC)		
500-word problem-solution essay	Rating scale	
Content	(0.0–4.5)	ea1c
Organization		ea1o
Language		ea1l
Referencing		ea1r
800-word discursive essay	Rating scale	
Content	(0.0–4.5)	ea2c
Organization		ea2o
Language		ea2l
Referencing		ea2e
Academic oral presentation	Rating scale	
Content	(0.0–4.5)	ea3c
Delivery		ea3d
Language		ea3l
Pronunciation/Fluency		ea3p
Online component	0.0–-1.0	eIndi
Overall course grade	Rating scale (0.0–4.5)	eover
Second English Course (AALC)		
600-word draft positive-argumentative essay	Rating scale (0.0–4.5)	
Content		aa1c
Organization		aa1o
Language		aa1l
Referencing		aa1r
Final Course Grade	Rating Scale (0-4.5)	
Derived Variable		Target variable
At-risk	0.0–3.5	"1"
Not at-risk	≥3.5	"0"

In each round of analysis, the derived binary variable (i.e., the final grade of the AALC) was the target variable. The component grades of all the assessments in the BALC and those of the first assessment in the AALC were the independent variables. The results of the online tasks in the BALC were also included as an independent variable. These variables are important, because the first English courses introduced concepts in basic academic literacy to students, and the BALC should have prepared students for the course in question, the AALC, as an advanced academic literacy course. Therefore, assessment components are appropriate predictors for the final grade of the AALC.

The cut-off point for the derived binary variable (i.e., the final grade of the AALC) was decided based on practical implications of the grade to the students. In the current study, students were considered to be "at-risk" if they attain a grade "B" or below. Based on Table 3, attaining a "B" grade is already being considered as "Good" and this was considered to be a sensible cut-off point at first. In practice, however, students need to achieve a better grade for

a better award for their degree. For example, in one of the departments in the research site < http://www.eie.polyu.edu.hk/docs/Programmes/Programme_Booklets/4year/42477/42477-BScIMT-1819-Aug2018.pdf>, attaining a grade point average of 3.2 of above is one of the factors for getting an Upper Second Class honor for their degree. Therefore, obtaining a B grade (i.e. 3.0) is not sufficient and grade B and below were considered as "at-risk".

Some independent variables have a direct mathematical relation with the dependent variable. The first assessment of the AALC was worth 20% of the overall grade, and this assessment had four assessment components aa1c (30%), aa1o (20%), aa1l (30%), and aa1r (20%). In other words, aa1c (30%), for example, was around 6% of the overall grade. The research team still saw the value of prediction, because it is practically meaningful to find out how important these components are for the final grade of the AALC. In addition, aalc and aall have equal importance mathematically, but it is still interesting to know which is statistically more important than the others. This practice is not new, as Jayaprakash et al. (2014) created a similar arrangement for their early alert system modeling.

Data Analysis

First, a classification tree analysis was conducted. All independent variables were included, and the computers decided which variables were to be retained. This study made use of the "rpart" library in R version 3.4.3, which is a classification and regression trees (CART) algorithm (Breiman, Friedman, Olshen, & Stone, 1984). Second, a logistic regression was conducted. The first trial of the logistic regression included all the independent variables, to identify the statistically significant predictors. The second and final trial of the logistics regression included only independent variables that were statistically significant predictors, and this model was treated as the finalized model. The estimates reported in the upcoming sections are from this final model.

Decision trees and logistic regression were used for their convenience and the availability of resources online. However, after reviewing the articles presented above, we believed that these two techniques were the easiest techniques to be applied by other practitioners.

Phase 2: Data Collection Procedures and Data Analysis

In Phase 2, participants attended a session at an annual event held at the research site. The author presented the results from Phase 1, including general information about the data-mining techniques and the data-mining results from each technique. Through interactions, most participants indicated that they did not know much about data mining and its techniques before the session. Participants were then invited to write down their general attitudes toward and perceptions of the data-mining techniques and the results. They were informed that their responses would be part of a research project, and would be published. The following questions were asked: (a) To what extent do you think learning analytics is useful/not useful? (b) Which technique do you think is the most useful/least useful? Why? (c) Within the field of learning analytics, what aspect(s) do you want to know more about? Thematic analysis was then adopted to analyze the responses of the participants.

Results

Phase 1 of this study considered the statistical strengths of two commonly used data-mining techniques: classification trees and logistics regression. Below are the results of analyzing the suitability these two techniques in a language-learning context.

Classification Tree Analysis (Phase 1)

The classification tree for this study attempted to identify at-risk students (i.e., receiving a grade "B" or below in the final grade for the AALC). Figure 1 shows this classification tree that uses the data set from two English Academic Purposes (EAP) courses, the BALC and the AALC. The four AALC component grades (content, organization, language, and referencing) were all present in the tree, helping to predict the students' outcomes. Among them, content was the most important predictor: In the training data set, 88% of students were classified as at-risk students if their content score was lower than 3.3. This is followed by the language score, as the second most important predictor. Other than the AALC components, the BALC organization score of the discursive essay was a predictor of student performance, but it affected less than 10% of the cases in the training data set.

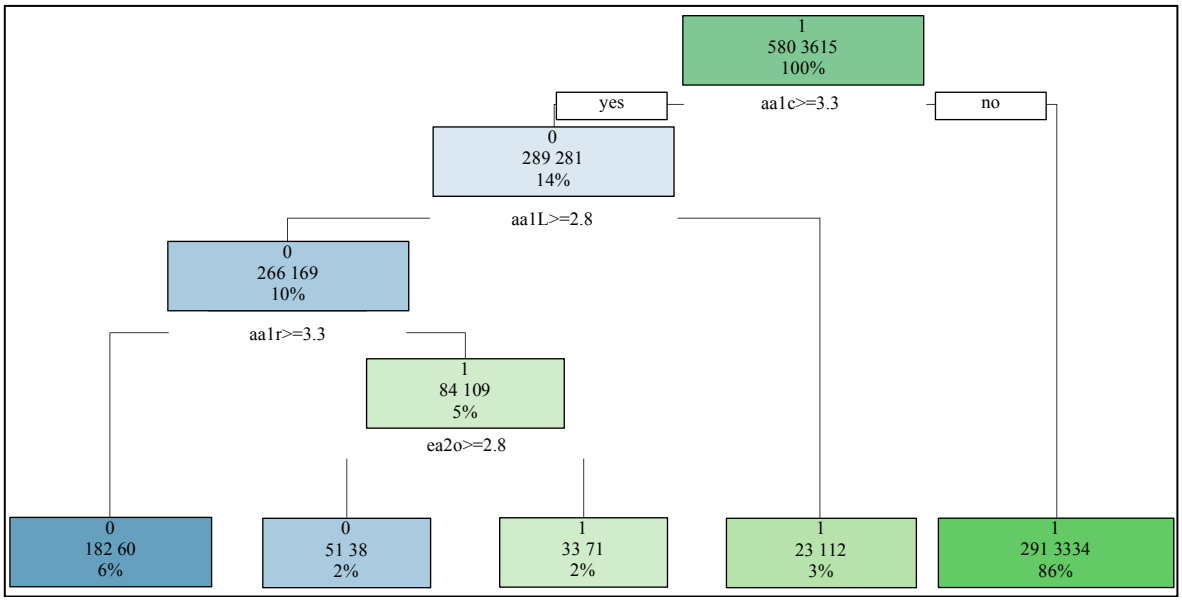

Figure 1. Classification tree.

Table 5 and Table 6 show the accuracy of the predictions made by the decision tree in the figure. The acronym used in this figure can be found in Table 4. The overall error rate of the training data set and testing data set was around 10%; in other words, in a typical class with 20 students in the research site, only 2 of the students would be wrongly classified, which can be considered acceptable from a practical perspective (note: there was no one common cut-off point for the accuracy rate). The false-negative rate was computed as well, at 2%, because it is important to see how many at-risk students were misclassified as not-at-risk students. With a low error rate, the results supported that classification trees are statistically appropriate for prediction purposes in the language-learning context.

Table 5
Misclassification in Decision Tree Analysis (Training Data Set)

PREDICTED	Actual		TOTAL
	Not At-risk	At-Risk	
Not At-risk	233	98	331
At-risk	347	351	3864
TOTAL	580	3615	4195

Error rate (training): (98 + 347) / 4,195 = 10.6%; false-negative: 98 / 3,615 = 2.7%.

Table 6
Misclassification in Decision Tree Analysis (Testing Data Set)

PREDICTED	Actual		TOTAL
	Not At-risk	At-Risk	
Not At-risk	103	33	136
At-risk	150	1512	1662
TOTAL	253	1545	1798

Error rate (testing): (33 + 150) / 1,798 = 10.2%; false-negative: 33 / 1,545 = 2.1%.

Logistics Regression

Logistics regression was run with the same batch of data described above. The logistic regression was to help identify at-risk students using all the predictors listed in Table 4. Within the data set, students who were at-risk (received a grade of "B" or lower) were marked as "1"; others not at-risk were marked as "0". Table 7 summarizes the statistically significant predictors. Unlike the classification tree analysis, more BALC components were retained in the final model, although, once again, all AALC components were retained as well.

Table 7 shows the coefficients or predictors identified in the logistics regression analysis. The odds ratio, also known as the risk ratio, can provide more insight into the importance of different predictors. The odds ratio refers to a change in the target variable when there is a one-unit change to the predictor; for example, a one-unit change in the referencing grade for the BALC problem-solution essay would lead to a (1.00 – 0.73) = 0.27 = 27% decrease in being at-risk. In other words, among the predictors retained, a change in the language component of the first AALC assessment would trigger the greatest change in the target variable (as it had the lowest odds ratio). Each unit of change in the language score would lead to a 74% decrease in being at-risk (see Table 7 below, for the odds ratio of the language score). The AALC content grade was second with the next greatest impact. With content and language as the components with the greatest impact, the logistics regression results were similar to those for the classification tree.

Table 7
Coefficients and Predictors for Logistics Regression

Predictors	Est.	*SE*	*OR*
(Intercept)	16.32	0.64	
BALC – Problem-solution essay: referencing	–0.32	0.09	0.73
BALC – discursive essay: organization	–0.47	0.11	0.63
BALC – Academic presentation: delivery	–0.38	0.10	0.68

AALC – Draft essay: content	–1.20	0.15	0.30
AALC – Draft essay: organization	–0.66	0.15	0.52
AALC – draft essay: language	–1.33	0.15	0.26
AALC – draft essay: referencing	–0.81	0.12	0.44

Note. *OR* = odds ratio.
*p < 0.01.

Table 8 and Table 9 show the error rates of the logistics regression presented above. The error rate of the logistics regression seemed to be comparable to that of the classification tree, around 10%. The false-negative rate was similar as well, at approximately 2%. In addition, the differences in the error rate between the training and testing data sets were minimal. All these suggest that logistics regression is an appropriate data-mining technique, in this context.

Table 8
Misclassification in Logistics Regression Analysis (Training Data Set)

	Actual		
PREDICTED	**Not At-risk**	**At-Risk**	**TOTAL**
Not At-risk	237	70	307
At-risk	343	3545	3888
TOTAL	580	3615	4195

Error rate (training): (343 + 70) / 4,195 = 10%; False-negative: 70 / 3,615 = 1.9%.

Table 9
Misclassification in Logistics Regression Analysis (Testing Data Set)

	Actual		
PREDICTED	**Not At-risk**	**At-Risk**	**TOTAL**
Not At-risk	100	35	135
At-risk	153	1510	1663
TOTAL	253	1545	1798

Error rate (testing): (153 + 35) / 1,798 = 10%; False-negative: 35 / 1,545 = 2.2%.

Phase 2: Teacher Questionnaire

Generally, the participating English teachers were a bit hesitant with data-mining techniques; only half of those who completed the questionnaire believed that the techniques are useful. Some presented doubts about the techniques themselves, such as the techniques "not being proved in our field," "teachers' qualitative judgement is more important in assuring [the accuracy] in language education," and "some predictions can be done without using these [data-mining techniques]." There were also doubts about these techniques' applicability in this context. For example, a respondent revealed that for data mining to work, grading needed to be consistent, and he was worried that some teachers were more lenient than others, thus affecting the data-mining results.

When teachers were asked which techniques seemed more useful, the only technique they reported was logistics regressions (n = 2), other than those who praised all techniques (n = 2). Many simply found the definitions or general concept of data mining useful. The teachers were also asked to suggest a possible direction for data mining. Many of those directions were

very traditional research questions featuring the relation between online tasks and final grades, or between performance in academic presentations and group discussions.

Discussion

This paper aimed to explore two questions: (a) which data-mining techniques can make a more accurate prediction of student performances in the language learning context? (b) Do language teachers have any preferences for the use of data-mining techniques?

Research Question 1

Interestingly, the results in this study cannot offer a definite answer to the first question, as the two techniques performed in a comparable manner with similar error and false-negative rates. Both methods had an accuracy rate of around 90%, which is considered to be good from a practical perspective. When these results are compared with those of Shahiri, Husaina, and Rashid (2015), the results were considered good. Therefore, there does not seem to be one method that is better than the others. As presented in the literature review, this result echoed Sarle (1994): that the techniques did not compete with each other. The accuracy and suitability depend on the data set and variables (Dreiseitl & Ohno-Machado, 2002). Therefore, both methods are accurate and acceptable for the present language-learning context.

It was also reassuring to see that the components identified in the techniques (all component grades from the AALC and one or two components of the BALC) were similar. The results showed that language and content development is important for students to achieve a good grade in the AALC. Shahiri et al. (2015) listed a number of important attributes used in predicting student performance, including students' cumulative grade point averages, internal assessment results, and student demographics. Therefore, the predictive power of the assessment components identified in this study is comparable that in other studies. It is only interesting to find that "internal assessment" should be individual assessment components (i.e., content grade for essay) instead of those in other studies (such quiz and exam). As the present study used only data from the learning management system, internal assessment results could be the most optimal predictors, among the good ones listed in the literature.

Research Question 2

Unfortunately, the language teachers did not seem very confident or interested in data-mining techniques. With few positive responses, they seemed to have concerns and doubts about the use of statistics in the language learning context. They may think that language is subjective, and thus, measured with a subjective system (e.g., essays, but not standardized exams). Wise and Vytasek (2017) identified numerous principles that are needed to better support teachers using analytics, including coordination, that is, making learning analytics an integral part of the educational context. The results of Phase 1 may have shown only the validity of the data-mining techniques, and teachers did not see how the techniques could be integrated into their context for taking action. This may be one reason why teachers are very positive about the results. Other this this, data mining seemed to be a new area for language teaching professionals. Bravo-Agapito, Bonilla, and Seoane (2018) conducted a review of the use of data mining in foreign language learning from 2012 to 2017. Factor analysis, as a traditional non-data-mining method, stood out as the most widely adopted method (44% of all studies in that review), far more than the data-mining techniques. Therefore, the low acceptance of data mining is not surprising.

Another important reason to explain the perception of teachers on these techniques may not be the fact that teachers have not yet experienced the techniques in full. When comparing the perceptions of teachers on different data mining techniques, one popular comment on classification tree would be the easy-to-interpret nature and actionable insights (Asif, Merceron, Ali, & Haider, 2017; Shahiri et al., 2015). Most teachers found this an important strength of classification tree because this makes it easy for teachers to explain the findings of the trees to students. However, the teachers did not get a chance to explain the results to a student in the current study, and it may be hard for them to imagine how easy it is to interpret the findings. The perceptions of teachers could have been different if they have tried to explain these results once.

Implications for Future Research

We cannot draw a definite conclusion about language teachers' perception of data mining. Further research is needed to understand teachers' perceptions, by involving teachers in using these data-mining techniques as part of the research in an educational context. In addition, this research presented only two algorithms to teachers, and more methods in the future can be presented to teachers for a more interesting comparison, including text mining. They will bring this line of research on data-mining techniques forward.

Limitations

This study has several limitations, and readers should interpret the results with caution. First, the accuracy of the predictions highly depends on the data set and how the variables are manipulated. In practice, the current study used "3.5" as the cut-off point due to its practical significance, and the predictive capability may change if this cut-off point is changed. The satisfactory results derived in this study may not be comparable to other studies. Second, the teachers' perceptions of data mining were affected by the results. Given the first limitation, teachers' perceptions may change if the results in another study are no longer satisfactory. In addition, this study examined only the perceptions of a small group of teachers at one research site. All these limitations may affect the generalizability of these results.

Conclusion

The results of this study suggest that commonly used data-mining techniques, such as classification trees and logistics regression, perform equally well in a language learning context in higher education. Both can be adopted in the language learning context to identify at-risk students. Similar to previous studies, internal assessment results and grades for content development and language remain the major predictors of students' performance. Despite good results in data-mining techniques, English-language teachers' attitudes regarding these techniques are concerning. Perhaps it is necessary to show them how these techniques can be integrated into an educational context, instead of the power of data mining only. Only if the process of such integration is considered to be effective and helpful, perception of frontline practitioners towards these techniques will not be changed. If they do not, it is hard for any instructional designers to develop any state-of-art systems. That is why the perspective of language teacher as the actual user is of ultimate importance and deserve further investigation.

References

Abdullah, N. H., Ramli, N. H. L., & Rafek, M. (2017). Mass lecture in language learning: What do the boys and girls think? *Journal of Advances in Humanities and Social Sciences*, *3*(2), 115–123.

Alfehaid, A. (2018). Integrating CALL with analytical rubrics for developing speaking skills. *CALL-EJ, 19*(2), 166–186.

Al-Maqaleh, B. M., & Abdullah, A. M. G. (2017). Intelligent predictive system using classification techniques for heart disease diagnosis. *International Journal of Computer Science Engineering (IJCSE)*, *6*(6), 145–151.

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, *113*, 177-194.

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, *4*(2), 17.

Bainbridge, J., Melitski, J., Zahradnik, A., Lauría, E. J., Jayaprakash, S., & Baron, J. (2015). Using learning analytics to predict at-risk students in online graduate public affairs and administration education. *Journal of Public Affairs Education*, *21*(2), 247–262.

Bravo-Agapito, J., Bonilla, C. F., & Seoane, I. (2018). Data mining in foreign language learning. *WIREs Data Mining Knowl Discov.* doi: https://doi.org/10.1002/widm.1287

Breiman, L., Friedman, J., Olshen, L., & Stone, J. (1984). *Classification and regression trees*. OH: Wadsworth.

Chingos, M. M., & Whitehurst, G. J. (2011). *Choosing blindly: Instructional materials, teacher effectiveness, and the common core*. Washington, DC: The Brookings Institution.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, *10*(1), 17–29.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, *35*(5–6), 352–359.

Finlay, J., Pears, R., & Connor, A. M. (2014). Data stream mining for predicting software build outcomes using source code metrics. *Information and Software Technology*, *56*(2), 183–198.

Goyal, M., & Vohra, R. (2012). Applications of data mining in higher education. *International Journal of Computer Science Issues (IJCSI)*, *9*(2), 113–120.

Llorente, R., & Morant, M. (2011). Data mining in higher education. In K. Funatsu (ed.). *New fundamental technologies in data mining*. IntechOpen. Retrieved from https://www.intechopen.com/download/pdf/13269

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Kongsuebchart, J., & Suppasetseree, S. (2018).The effect of a weblog-based electronic portfolio on Thai EFL undergraduate students' English writing skills. *CALL-EJ, 19*(2), 28–46.

Kwok, S. (2016, May 10). How Hong Kong universities can give students the best start in life and help fix the education system. *South China Morning Post.* Retrieved from https://www.scmp.com/comment/insight-opinion/article/1942610/how-hong-kong-universities-can-give-students-best-start-life

Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, *36*(1), 2–17.

Park, Y., Yu, J. H., & Jo, I. H. (2016). Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education, 29,* 1–11.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics, 40*(6), 601–618. doi:10.1109/TSMCC.2010.2053532

Sahoo, G. (2013). Study of parametric performance evaluation of machine learning and statistical classifiers. *International Journal of Information Technology and Computer Science, 5*, 57–64.

Sarle, W. S. (1994). Neural networks and statistical models. *Proceedings of the 19th annual SAS users group international conference*.

Shahiri, A. M., Husain, W., Rashida, N. A. (2015). A review on predicting students' performance using data mining techniques. *Procedia Computer Science*, *72*, 414–422.

Tan, P., Strinbach, M., & Kumar, V. (2013). *Introduction to data mining*. Essex, UK: England: Pearson.

Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. In C. Lang, G. Siemens, A. Wise, & D. Gašević. (eds.). *Handbook of learning analytics (pp 151-160)*, Society for Learning analytics Research. Retrieved from https://www.researchgate.net/profile/Helene_Fournier/publication/325070028_A_Critical_Perspective_on_Learning_Analytics_and_Educational_Data_Mining/links/5af49b5ea6fdcc0c030af54d/A-Critical-Perspective-on-Learning-Analytics-and-Educational-Data-Mining.pdf#page=151