Computers in Human Behavior 107 (2020) 105584

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full Length Article

Predicting at-risk university students in a virtual learning environment via a machine learning algorithm

Kwok Tai Chui[a,*], Dennis Chun Lok Fung[b], Miltiadis D. Lytras[c,d], Tin Miu Lam[b]

[a] *Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China*
[b] *Faculty of Education, The University of Hong Kong, Hong Kong, China*
[c] *Management Information Systems Department, The American College of Greece, Athens, Greece*
[d] *Distinguished Scientist, King Abdulziz University, Jeddah, Saudi Arabia*

ARTICLE INFO

Keywords:
Academic performance
At-risk students
Event prediction
Higher education
Machine learning
Virtual learning environments

ABSTRACT

A university education is widely considered essential to social advancement. Ensuring students pass their courses and graduate on time have thus become issues of concern. This paper proposes a reduced training vector-based support vector machine (RTV-SVM) capable of predicting at-risk and marginal students. It also removes redundant training vectors to reduce the training time and support vectors. To examine the effectiveness of the proposed RTV-SVM, 32,593 university students on seven courses were chosen for performance evaluation. Analysis reveals that the RTV-SVM achieved a training vector reduction of at least 59.7% without altering the margin or accuracy of the classifier. Moreover, the results showed the proposed method to be capable of achieving overall accuracy of 92.2–93.8% and 91.3–93.5% in predicting at-risk and marginal students, respectively.

1. Introduction

Data analytics (Lytras, Raghavan, & Damiani, 2017; Yao et al., 2017) and learning analytics (Toetenel & Rienties, 2016; Wilson, Watson, Thompson, Drew, & Doyle, 2017) are emerging research fields that have arisen in the wake of the tremendous advances in artificial intelligence and data mining achieved in the past few decades. Higher education is important because students study and equip advanced technical skills to contribute the society. Whilst millions of students graduate from universities worldwide every year, there are many who fail their courses and are forced to retake them before progressing to the next year of study or to graduation. Evidence suggests that many of these students lose confidence and suffer from depression (Hill, Taroslavsky, & Pettit, 2015; Schöne, Tandler, & Stiensmeier-Pelster, 2015) and, in the worst-case scenario, defer their studies or withdraw from university, which significantly increases both the workload of academic staff and university expenditure (Cemalcilar & Gökşen, 2014). From a global perspective, if a considerable number of students leave the university due to academic failure, not only will the reputation of universities themselves suffer, but societal aspirations will also be undermined (Letseka & Maile, 2008). As a result, there is a pressing need to develop an accurate means of predicting academically at-risk university students to minimise the potential waste of public funds spent on students who subsequently fail (Jia & Maloney, 2015; Ogunyemi, 2005).

The current study adopted SVM, which is an influential data classification method used in a range of areas, including imaging (Nandi, Srivastava, & Shah, 2017; Taravat, Frate, Cornaro, & Vergari, 2015), bioinformatics (Chui, Tsang, Chi, Ling, & Wu, 2016; Chui et al., 2015), energy management (Chia, Lee, Shafiabady, & Isa, 2015; Meng, Luo, & Gao, 2016) and fault detection (Liu, Yang, Zhang, Wang, & Chen, 2016; Zheng, Pan, & Cheng, 2017). Existing SVM applications can be generally divided into two groups: algebraic view and geometric view. The former's aim is to minimise classification errors and reduce computational costs. Applications falling into this group include soft margin SVM (Dias & Neto, 2017), kernel SVM (Kirar & Agrawal, 2017), support vector regression (Okujeni, van der Linden, Suess, & Hostert, 2017), ν-SVM (Gu & Sheng, 2017) and sequential minimal optimisation (SMO; Pham, Bui, Prakash, Nguyen, & Dholakia, 2017). The latter group includes dual representation SVM (Gotoh & Uryasev, 2017) and convex hull SVM (Sun, Zhang, Wang, Ren, & Jin, 2016).

In many engineering problems, researchers adopt SVM and treat it as a black box for data classification. The main task is to construct the feature vector and evaluate the performance of the classifier using traditional built-in kernel functions such as the linear kernel, radial basis kernel, polynomial kernel and sigmoid kernel. However, for more

* Corresponding author.
E-mail addresses: ktchui3-c@my.cityu.edu.hk (K.T. Chui), clfung@hku.hk (D.C.L. Fung), mlytras@acg.edu (M.D. Lytras), tinmiu@connect.hku.hk (T.M. Lam).

https://doi.org/10.1016/j.chb.2018.06.032
Received 19 December 2017; Accepted 25 June 2018
Available online 27 June 2018
0747-5632/ © 2018 Elsevier Ltd. All rights reserved.

complex problems, traditional kernel functions may be unable to achieve the desired performance, and a pre-computed or customised kernel is thus required.

Furthermore, when the number of training vectors is large (as is frequently the case today, given the large datasets researchers often deal with), the question of how to improve the efficiency of SVM while retaining classification accuracy, is worth answering. As support vectors are essential training vectors in determining the hyperplane of an SVM classifier, two relevant properties must be considered. First, the distance between a training vector and hyperplane is greater than that between a support vector and hyperplane. Second, most support vectors are located at the boundary of their respective class. This paper proposes a reduced training vector-based SVM (RTV-SVM) designed to remove redundant training vectors with the aim of reducing the training time and preserving the support vectors (without scarifying the classification accuracy).

The contributions of this paper are (i) Reduction of training vector and thus training time; (ii) Preserving the support vectors in order to avoid deteriorating in classification accuracy; (iii) Predicting both the marginal and at-risk students.

The remainder of the paper is organised as follows. Literature review is presented in Section 2. Section 3 discusses the RTV-SVM methodology. Its application to predicting at-risk university students is then evaluated in Section 4. Section 5 reports and discusses a performance comparison between the RTV-SVM and the methods adopted in related work, and Section 6 concludes the paper with a summary.

2. Literature review

Data processing, visualization, educational data mining, technology learning enhancement and predictive models are tools of learning analytics that aims at providing meaningful actions (Scheffel, Drachsler, Stoyanov, & Specht, 2014). Many works have shown the advantages of using learning analytics. It is worth having customized model that fits the specific application (need of learners) and thus it can achieve the biggest impact on them, for instance, to intervene at-risk students or provide instructional content and feedback.

Trivial solutions for dealing with the issue of at-risk students include (i) replacing tests and examinations with attendance as the course assessment method, allowing students to pass their courses merely by showing up and thus achieving a good attendance record, or (ii) lowering the passing score for course examinations to allow underachievers to pass. Obviously, neither method is a sensible solution because both would threaten the quality of university graduates. To find an alternative, several recent research projects have investigated a variety of approaches to assisting at-risk students academically (e.g. Gibbs & Culleiton, 2016; Lindt & Blair, 2017).

Although the first step to improving the likelihood of academic success is the identification of at-risk students as early as possible, researchers have also drawn attention to the various analytical methods used to predict such students. The various means of prediction models explored in this body of literature include (i) forecasting which students will fail their courses (Gašević, Dawson, Rogers, & Gasevic, 2016; Hu, Lo, & Shih, 2014; Marbouti, Diefes-Dux, & Madhavan, 2016); (ii) identifying students likely to defer their studies or fail to finish them on time (Jia & Maloney, 2015; Lakkaraju et al., 2015); (iii) predicting which students will drop out (Duarte, Ramos-Pires, & Gonçalves, 2014); and (iv) identifying students who will graduate with a narrow pass (Alharbi, Cornford, Dolder, & De La Iglesia, 2016). The study reported in this paper explored category (i) because it is considered to be the most fundamental stage and one that leads to the other categories. Moreover, the situations in categories (ii)-(iv) were also deemed to be more complex, as they may be attributed to multiple issues and factors. In terms of the mechanisms of prediction in the studies cited above, various machine learning algorithms, including the random forest, logistic regression, support vector machine (SVM), decision tree, C4.5, CART, multiple linear regression, multi-layer perception, naïve-based classifier, K-nearest neighbor, maximum likelihood probit and auto-classifier algorithms, were employed.

Typical challenges of learning analytics include (i) big data framework: In order to handle massive amounts of data and support heavy querying, the monitoring system should possess a robust big data framework. The number of sensors and users keep increasing which desire good scalability and flexibility; (ii) data collection: This is the second fundamental issues that have to be addressed when it comes to learning analytics. Problems are what kinds of data to be collected and how many samples require. The availability of big data storage allows new system to implement in a way to collect as more features and samples as possible. More features can provide better understanding of the problems and hidden pattern might be concluded. More samples can ensure a more generic model because it can generalize the situation. It is often that many schools adopt small class teaching scheme in which the sample size could be low. For classification problem, one-class classification algorithm can utilize to solve the concern in the lack of sample in certain class label; (iii) data analysis: It is related to missing data handling and data pre-processing. For example, an instructor may create a student profile to isolate an assignment that requires grading, test the ease of submission process, or to determine if there are any gaps in the presentation of the curriculum as it appears for students. Creation of a non-existent learner introduces redundant information that appears in the course without identification. This data does not represent student information but rather misinformation created by the instructor that flows into the big data pool of information; (iv) privacy: Data privacy is essential element which contains confidential information, for instance, name, gender, age, identity and grade. The system should share information for big data analytics that is needed to provide high-quality learning analytics to students and teachers, while assuring privacy simultaneously; (v) security: Data security is key to retain and gain users trust on privacy. The monitoring system consists of massive amounts of interconnected devices. Many of the existing devices are lack of security because of technical challenges to redesign traditional hardware and software. Blackmailer, disgruntled person, terrorist and hacker can access the data of the devices via any single device (Marin, Pawlowski, & Jara, 2015), cloud infrastructure (Stergiou, Psannis, Kim, & Gupta, 2018) or network (Kaur, Kaur, & Sood, 2017); and (vi) algorithms: It is usually a tedious process to obtain the best algorithm because there are too many existing algorithms. We have to evaluate the performance using some algorithms and if it is not good enough, algorithm enhancement, deep learning or ensemble learning can be used. These are usual difficulties in data science and analytics.

3. RTV-SVM methodology

The RTV-SVM methodology is divided into four parts in this section: (i) a basic definition of inputs; (ii) the tier-1 elimination of training vectors via the multivariable normal approach; (ii) the tier-2 elimination of training vectors via vector transformation; and (iv) building an SVM classifier using SMO. Binary classification is also considered.

3.1. Basic definition of RTV-SVM inputs

Define a set of N training feature vectors $X_{train} = \{x_1, x_2, ..., x_N\}$ with corresponding labels $Y_{train} = \{y_1, y_2, ..., y_N\}$, where $y_i \in \{-1, +1\}$. The objectives of the RTV-SVM are to (i) eliminate redundant training vectors; and (ii) train the SVM classifier for data classification. To obtain a valid formulation, the RTV-SVM deals with a convex hull for the training vectors in each class, and the classification problem is separable.

3.2. Tier-1 elimination of training vectors

The multivariate normal distribution of the X_{train} of each class can be found using

$$f(x|\mu, \sigma^2) = (\mu, D), \tag{1}$$

$$\mu = (\sum_{i=1}^{N} x_i)/N \text{ and } \sigma^2 = (\sum_{i=1}^{N} (x_i - \mu)^2)/N, \tag{2}$$

where μ is the mean of the normal model $f(x|\mu, \sigma^2)$, D is the MxM diagonal covariance matrix, with σ^2on its diagonal, and m is the number of dimensions. The training vector with respect to the multivariate normal probability distribution function is given by

$$P(x) = e^{(-0.5(x-\mu)^T D^{-1}(x-\mu))}/(2\pi)^{d/2}\sqrt{|D|}. \tag{3}$$

Based on equations (1)–(3), it can be seen that the vectors close to the boundary have a smaller probability than those close to the centre of the normal distribution. The RTV-SVM obtains N_t training vectors in each class with the smallest probability. The value of N_t is determined automatically, as detailed below.

The average probability of P(x) in one of the cardinalities of class C_{train}, $|C_{train}|$, and can be calculated by

$$\overline{P(x)} = \Big[\sum_{i=1}^{|C_{train}|} P(x_i)\Big]/|C_{train}|. \tag{4}$$

The minimum and maximum average probabilities of P(x) are found using

$$\overline{P(x)}_{\min} = \Big[\sum_{j=1}^{|C_{\min}|} P(x_i)\Big]/|C_{\min}| \text{ and } \overline{P(x)}_{\max} = \Big[\sum_{k=1}^{|C_{\max}|} P(x_i)\Big]/|C_{\max}| \tag{5}$$

$$\begin{array}{l} x_j \in C_{\min} = \{x_i | P(x_i) < \overline{P(x)}, x_i \in C_{train}\} \\ x_k \in C_{\max} = \{x_i | P(x_i) \geq \overline{P(x)}, x_i \in C_{train}\} \end{array}. \tag{6}$$

Here, we define the ratio of minimum average probability to maximum average probability as

$$R_{\max}^{\min} = \overline{P(x)}_{\min} \Big/ \overline{P(x)}_{\max}. \tag{7}$$

This ratio is an indicator of the distribution of training vectors with respect to the decision boundary and centre of the class. A small value indicates that most x_i are close to the centre of the class, whereas a large value suggests that most are close to the decision boundary. N_t is determined by $R_{\max}^{\min}$. Next, define $R_{\max}^{\min} \geq R_{th} = 0.35$, where R_{th} is the threshold. N_t is calculated by

$$N_t = \begin{cases} N_t|_{class}, & R_{\max}^{\min} = R_{th} \\ \Big\lceil \frac{\overline{P(x)} - \overline{P(x)}_{\min}}{\overline{P(x)}_{\max} - \overline{P(x)}_{\min}} \bullet |C_{train}| \Big\rceil, & R_{\max}^{\min} \geq R_{th} \end{cases}, \tag{8}$$

where $N_t|_{class}$ denotes the number of x_i in the class.

3.3. Tier-2 elimination of training vectors

The number of training vectors is further reduced by vector transformation. In binary classification, we define $X_{0,train}$ and $X_{1,train}$ as the training vectors for Classes 0 and 1, respectively:

$$X_{train} = X_{0,train} \cup X_{1,train} \tag{9}$$

$$X_{0,train} = \{x_{0,1}, x_{0,2}, ..., x_{0,N0}\} \text{ and } X_{1,train} = \{x_{1,1}, x_{1,2}, ..., x_{1,N1}\} \tag{10}$$

The origins of the coordinate system of $X_{0,train}$ and $X_{1,train}$ are transformed into centres μ_{x0} and μ_{x1}, respectively:

$$x'_{0,i} = x_{0,i} - \mu_{x0},\ i \in [1, N0] \text{ and } x'_{1,j} = x_{1,j} - \mu_{x1},\ j \in [1, N1] \tag{11}$$

Now define the vector joining $\mu_{x0}$and μ_{x1} as

$$\mu'_{x1} = \mu_{x1} - \mu_{x0} \text{ and } \mu'_{x0} = \mu_{x0} - \mu_{x1} \tag{12}$$

The RTV-SVM projects $x_{0,i}$', $i \in [1, N0]$ onto μ'_{x1}, and $x_{1,j}$', $j \in [1, N1]$ onto μ'_{x0}, giving us

$$|\mu'_{x1}| \cos\alpha_i = \frac{x'_{0,i} \bullet \mu'_{x1}}{|x_{0,i}'|} \text{ and } |\mu'_{x0}| \cos\beta_j = \frac{x'_{1,j} \bullet \mu'_{x0}}{|x'_{1,j}|}, \tag{13}$$

where α_i is the angle between $x'_{0,i}$ and μ'_{x1}, and β_j is the angle between $x'_{1,j}$and μ'_{x0}.Putting (11) and (12) into (13), we have

$$|\mu_{x1} - \mu_{x0}| \cos\alpha_i = \frac{(x_{0,i} - \mu_{x0}) \bullet (\mu_{x1} - \mu_{x0})}{|x_{0,i} - \mu_{x0}|} \tag{14}$$

$$|\mu_{x0} - \mu_{x1}| \cos\beta_j = \frac{(x_{1,j} - \mu_{x1}) \bullet (\mu_{x0} - \mu_{x1})}{|x_{1,j} - \mu_{x1}|} \tag{15}$$

$$\delta(|\mu_{x1} - \mu_{x0}| \cos\alpha_i) = \begin{cases} 1, & |\mu_{x1} - \mu_{x0} \cos\alpha_i| \geq 0 \\ 0, & otherwise \end{cases} \tag{16}$$

$$\delta(|\mu_{x0} - \mu_{x1}| \cos\beta_j) = \begin{cases} 1, & |\mu_{x0} - \mu_{x1} \cos\beta_j| \geq 0 \\ 0, & otherwise \end{cases} \tag{17}$$

In (16) and (17), the value of 1 indicates that the training vectors in Classes 0 and 1, respectively, are preserved. The training vector in Class 0 is eliminated if it satisfies the three following conditions.

$$\begin{array}{l} C1_{C0}: \delta(|\mu_{x1} - \mu_{x0}| \cos\alpha_i) = 0 \\ C2_{C0}: I_{0,i} \neq \underset{i \in [1,N0], \cos\alpha_i \geq 0}{\arg\min} \cos\alpha_i \\ C3_{C0}: I_{0,i} \neq \underset{i \in [1,N0], \cos\alpha_i < 0}{\arg\min} -\cos\alpha_i \end{array}, \tag{18}$$

where $I_{0,i}$ is the index of the $X_{0,i}$ training vector. Similarly, the training vector in Class 1 is eliminated when

$$\begin{array}{l} C1_{C1}: \delta(|\mu_{x0} - \mu_{x1}| \cos\beta_j) = 0 \\ C2_{C1}: I_{1,j} \neq \underset{j \in [1,N1], \cos\beta_j \geq 0}{\arg\min} \cos\beta_j \\ C3_{C1}: I_{1,j} \neq \underset{i \in [1,N1], \cos\beta_j < 0}{\arg\min} -\cos\beta_j \end{array}, \tag{19}$$

where $I_{1,j}$ is the index of the $X_{1,j}$ training vector.

3.4. Construction of SVM classifier

After eliminating the training vectors in both classes, the remaining training vectors are denoted as X_{remain}. The RTV-SVM performs SMO to calculate discriminant function $\phi(x) = \omega^T x + b$ using X_{remain}. The quadratic optimisation problem is defined as

$$(\omega*, b*) = \underset{\omega,b}{\arg\min} \frac{1}{2}\|\omega\|^2 \text{s.t.} y_i(\omega^T x_i + b) \geq 1,\ x_i \in X_{remain},\ y_i \in \{-1, +1\}. \tag{20}$$

Using the Karush-Kuhn-Tucker condition (Bertsekas, 1999), we can formulate the problem using the Lagrange function:

$$L(\omega, b, \chi) = \frac{1}{2}\omega^T\omega - \sum_{i=1}^{N'} \chi_i [y_i(\omega^T x + b)] - 1, \tag{21}$$

where $L(\omega, b, \chi)$ is the Langrage function, $\chi = \{\chi_1, \chi_2, ..\chi_{N'}\}$ denotes the vector of Lagrange multipliers, and $i \in \{1, ..., N'\}$ denotes the cardinality of X_{remain}.

4. Predicting at-risk university students using RTV-SVM method

This section first discusses the dataset for predicting at-risk university students, and the proposed RTV-SVM is then applied to that dataset and evaluated in four scenarios: (i) no reduction in training vectors; (ii) tier-1 elimination alone; (iii) tier-2 elimination alone; and (iv) both tier-1 and tier-2 elimination. The evaluation criteria include the percentage decrease in training vectors, the time taken to train and test the classifier, and the sensitivity, specificity and overall accuracy of the classifier.

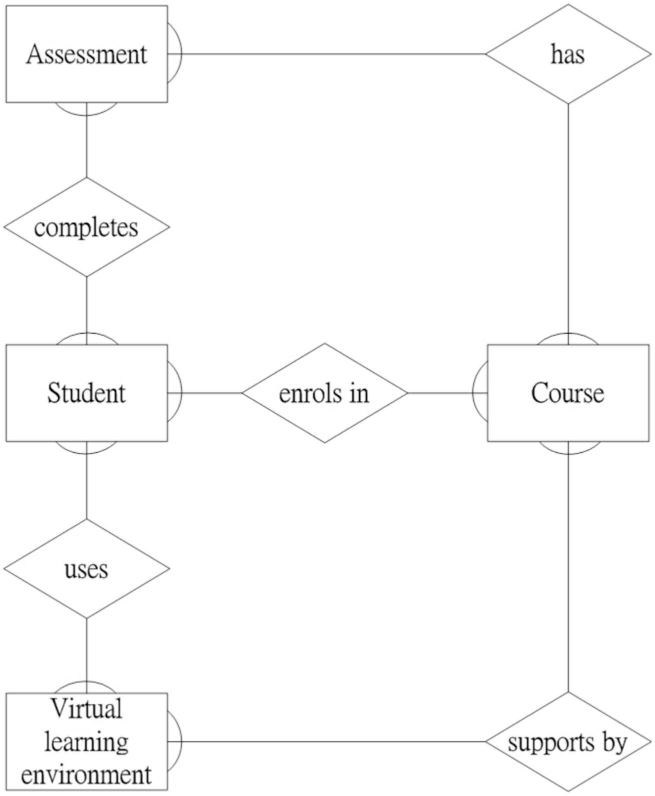

Fig. 1. Entity-relationship model for student, virtual learning environment, course and assessment in open university.

4.1. OULA dataset

The Open University Learning Analytics (OULA) dataset (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015), which aim is to provide an early prediction of at-risk students based on their demographic data and interaction with a virtual learning environment, was retrieved for the study reported herein. The OULA dataset was collected by the largest university in the UK, i.e. the Open University (OU), in the 2013-14 academic year and comprises seven courses, 32,593 students (17,875 men and 14,718 women) and four semesters. Failure and distinction are defined as scores below 40 and greater than or equal to 85, respectively. The distribution of student marks is (i) Fail: 7052 students; (ii) Pass: 12,361 (Distinction: 3024); and (iii) Withdraw: 10,156 (early drop or withdrawal before two weeks: 4474). Fig. 1 illustrates the relationship amongst the seven comma-separated values (CSV) files (e.g. assessment.csv and studentInfo.csv) in the OULA dataset. The files are divided into three domains, student activities, student demographics and module presentation.

The OU grading scale for the OULA dataset is presented in Table 1 (Assessment Handbook for Undergraduate and Postgraduate Module, 2017). Performance evaluation in the dataset consists of two parts. In the first part (see Section 3.2 below), the goal is to predict at-risk students who will fail their course (i.e. receive a grade of E, F or G) using the RTV-SVM for binary classification. In the second part (i.e. Section 3.3), the aim is to predict at-risk students (grade E, F or G) and marginal students (grade D) using the RTV-SVM for multi-class classification.

Table 1
Grading scale of the Open University (UK).

Grade	University scale score	Performance standard
A	85–100	Pass 1 (Distinction)
B	70–84	Pass 2
C	55–69	Pass 3
D	40–54	Pass 4
E	30–39	Bare fail
F	15–29	Fail
G	0–14	Bad fail

Table 2
Binary classification dataset used to predict at-risk students via RTV-SVM.

Course	Number of students in Class 0 (All, male, female)	Number of students in Class 1 (All, male, female)
AAA	(487,292,195)	(91,52,39)
BBB	(3077,349,2728)	(1767,237,1530)
CCC	(1180,869,311)	(781,636,145)
DDD	(2227,1334,893)	(1412,904,508)
EEE	(1294,1142,152)	(562,514,48)
FFF	(2978,2438,540)	(1711,1459,252)
GGG	(1118,202,916)	(728,147,581)

4.2. Performance evaluation of RTV-SVM classifier: predicting at-risk students

Two classes were defined for the binary classification problem used to test the RTV-SVM's performance, namely, Class 0: Pass; and Class 1: Fail. Thus, the RTV-SVM makes a decision on either Class 0 or Class 1. The binary classification dataset is provided in Table 2, which also gives the total number of students and the number of male and female students in each class. Note that the course names are anonymous.

The RTV-SVM was applied to each course using five-fold cross-validation. To determine the effectiveness of tier-1 and tier-2 elimination, four scenarios were considered: S1: no reduction in training vectors; S2: tier-1 elimination alone; S3: tier-2 elimination alone; and S4: both tier-1 and tier-2 elimination. Seven criteria, namely, the percentage reduction in training vectors (R_{tv}), margin variation with respect to S1 (M_r), training time (T_1), testing time (T_2), sensitivity (S_e), specificity (S_p) and overall accuracy (OA), were selected to evaluate the RTV-SVM classifier. The results are presented in Table 3, with the most notable singled out for analysis below.

(1) Criterion R_{tv} shows that tier-1 and tier-2 elimination can effectively reduce the number of training vectors without altering the margin or the support vectors of the classifier (M_r remains unchanged) and thus retaining the accuracy. The S4 results suggest that both tier-1 and tier-2 elimination should be applied to maximise the reduction of training vectors.
(2) T_1 reduces significantly with a decrease in R_t, which implies that reducing the number of training vectors is important to reducing the training time for a classifier. The T_1 reduction becomes even more significant with an increase in the size of the dataset. It is worth mentioning that T_2 remained unchanged throughout S1−S4 (other than a negligibly small fluctuation) because the number of testing datasets was identical in all four scenarios.
(3) When it comes to S_e, S_p and OA, it can be seen that the proposed method yields good performance of at least 92.1%, 92.0%, and 92.2%, respectively. The values of these criteria are equal in S1−S4 because M_r remains unchanged.

4.3. Performance evaluation of RTV-SVM classifier: predicting at-risk and marginal students

Consider a multi-class classification problem using the RTV-SVM, three classes are defined, namely, Class 0: Pass; Class 1: Marginal; and Class 2: Fail. In the current study, 1-against-1 multi-class SVM with N (N-1)/2 = 3 classifiers, where N = 3 is the number of training classes, was chosen because it generally achieves a higher degree of accuracy and requires less training and testing time (Hsu & Lin, 2002). The

Table 3
Performance for binary classification problem: predicting at-risk students via RTV-SVM.

Course	Scenario	Criteria for evaluation of RTV-SVM classifier						
		R_{tv} (%)	M_r	T_1 (s)	T_2 (s)	S_e	S_p	OA
AAA	S1	0	N/A	0.721	0.151	0.921	0.927	0.924
	S2	39.3	No	0.413	0.149	0.921	0.927	0.924
	S3	20.6	No	0.548	0.154	0.921	0.927	0.924
	S4	59.9	No	0.252	0.153	0.921	0.927	0.924
BBB	S1	0	N/A	4.52	0.893	0.940	0.936	0.938
	S2	38.8	No	2.67	0.896	0.940	0.936	0.938
	S3	34.5	No	3.05	0.894	0.940	0.936	0.938
	S4	73.3	No	1.09	0.899	0.940	0.936	0.938
CCC	S1	0	N/A	1.95	0.368	0.929	0.921	0.925
	S2	34.6	No	1.21	0.376	0.929	0.921	0.925
	S3	33.9	No	1.34	0.374	0.929	0.921	0.925
	S4	68.5	No	0.64	0.371	0.929	0.921	0.925
DDD	S1	0	N/A	3.87	0.702	0.931	0.933	0.932
	S2	29.8	No	2.63	0.697	0.931	0.933	0.932
	S3	31.5	No	2.36	0.699	0.931	0.933	0.932
	S4	61.3	No	1.35	0.692	0.931	0.933	0.932
EEE	S1	0	N/A	1.92	0.337	0.924	0.928	0.926
	S2	32.5	No	1.27	0.346	0.924	0.928	0.926
	S3	27.4	No	1.36	0.341	0.924	0.928	0.926
	S4	59.9	No	1.06	0.344	0.924	0.928	0.926
FFF	S1	0	N/A	4.95	0.864	0.924	0.920	0.922
	S2	33.1	No	3.17	0.873	0.924	0.920	0.922
	S3	34.6	No	2.98	0.862	0.924	0.920	0.922
	S4	67.7	No	1.44	0.871	0.924	0.920	0.922
GGG	S1	0	N/A	1.97	0.335	0.927	0.933	0.930
	S2	47.2	No	0.96	0.339	0.927	0.933	0.930
	S3	24.6	No	1.46	0.338	0.927	0.933	0.930
	S4	71.8	No	0.468	0.340	0.927	0.933	0.930

Table 4
Dataset for multi-class classification problem predicting at-risk and marginal students via the RTV-SVM.

Course	Number of students in Class 0 (All, male, female)	Number of students in Class 1 (All, male, female)	Number of students in Class 2 (All, male, female)
AAA	(441,263,178)	(46,29,17)	(91,52,39)
BBB	(2701,303,2398)	(376,46,330)	(1767,237,1530)
CCC	(1171,864,307)	(9,5,4)	(781,636,145)
DDD	(1909,1153,756)	(318,181,137)	(1412,904,508)
EEE	(548,479,69)	(746,663,83)	(562,514,48)
FFF	(1762,1452,310)	(1216,986,230)	(1711,1459,252)
GGG	(1002,180,822)	(116,22,94)	(728,147,581)

dataset for this problem, which is similar to that in Table 2, is described in Table 4, which also gives the total number of students and number of male and female students in each class.

Similar to the case in Section 3.2, the RTV-SVM was applied to each of the courses using five-fold cross-validation. However, because the sample size of Class 1 in CCC was too small, cross-validation was not considered for that course. The results for the seven criteria are displayed in Table 5, and their analysis is summarised below.

(1) Similar to the binary classification problem, criterion R_{tv} shows that tier-1 and tier-2 elimination can effectively reduce the number of training vectors with an identical M_r. Thus, the reduction of training vectors will not scarifying the accuracy of the classifier. Adopted both tier-1 and tier-2 elimination can thus maximise the method's performance.
(2) T_1 also reduces significantly with increases in R_{tv}. It can be seen that T_1 in Table 5 is much larger than that in Table 3 because the required number of classifiers increases from 1 to 3. T_2 is also larger for the same reason, as more time is needed to check the class label of the testing data. Also, T_2 holds steady across S1−S4.

Table 5
Performance for binary classification problem: predicting both the at-risk and marginal students via RTV-SVM.

Course	Scenario	Criteria for evaluation of RTV-SVM classifier						
		R_{tv} (%)	M_r	T_1 (s)	T_2 (s)	S_e	S_p	OA
AAA	S1	0	N/A	2.01	0.386	0.917	0.921	0.919
	S2	38.5	No	1.21	0.388	0.917	0.921	0.919
	S3	21.2	No	1.53	0.389	0.917	0.921	0.919
	S4	59.7	No	0.744	0.388	0.917	0.921	0.919
BBB	S1	0	N/A	10.2	1.942	0.937	0.933	0.935
	S2	38.2	No	6.12	1.945	0.937	0.933	0.935
	S3	34.6	No	6.98	1.938	0.937	0.933	0.935
	S4	72.8	No	2.49	1.939	0.937	0.933	0.935
CCC	S1	0	N/A	5.11	1.02	0.919	0.915	0.917
	S2	34.7	No	3.27	1.03	0.919	0.915	0.917
	S3	34.0	No	3.45	1.01	0.919	0.915	0.917
	S4	68.7	No	1.51	1.02	0.919	0.915	0.917
DDD	S1	0	N/A	10.4	1.75	0.928	0.924	0.926
	S2	30.1	No	7.07	1.78	0.928	0.924	0.926
	S3	31.3	No	6.78	1.77	0.928	0.924	0.926
	S4	61.4	No	3.80	1.79	0.928	0.924	0.926
EEE	S1	0	N/A	5.01	0.874	0.918	0.924	0.921
	S2	35.8	No	3.11	0.877	0.918	0.924	0.921
	S3	26.1	No	3.58	0.876	0.918	0.924	0.921
	S4	61.9	No	1.75	0.878	0.918	0.924	0.921
FFF	S1	0	N/A	14.2	2.03	0.916	0.910	0.913
	S2	32.6	No	9.37	2.04	0.916	0.910	0.913
	S3	33.9	No	8.86	2.03	0.916	0.910	0.913
	S4	66.5	No	4.41	2.02	0.916	0.910	0.913
GGG	S1	0	N/A	5.25	0.872	0.925	0.921	0.923
	S2	49.7	No	2.47	0.873	0.925	0.921	0.923
	S3	25.2	No	3.83	0.875	0.925	0.921	0.923
	S4	74.9	No	1.19	0.871	0.925	0.921	0.923

(3) The S_e, S_p and OA values in the multi-class classification problem are good, achieving at least 91.6%, 91.0% and 91.3%, respectively. The slight deterioration in performance is attributable to the increasing complexity of the problem (three classifiers). The values of these criteria are the same in S1−S4 because M_r remains unchanged.

5. Performance comparison between RTV-SVM and related methods

As noted in the Introduction and Literature Review sections of the paper, predictions of at-risk students include predictions of (i) students who will fail their courses (Gašević et al., 2016; Hu et al., 2014; Marbouti et al., 2016); (ii) students who will defer their studies or fail to finish them on time (Jia & Maloney, 2015; Lakkaraju et al., 2015); (iii) students who will drop out (Duarte et al., 2014); and (iv) students who will graduate with a narrow pass (Alharbi et al., 2016). The study reported herein considered category (i) alone.

Table 6 presents a performance comparison between the RTV-SVM and the methods used in related studies considering categories (i)-(iv). With regard to category (i), it can be seen that the proposed method achieved significantly better performance than the methods reported by Gašević et al. (2016) and Marbouti et al. (2016). Compared with the method adopted by Hu et al. (2014), the RTV-SVM yielded a 3.5–5.1% lower degree of OA. However, their work may not offer a reliable comparison because it did not involve cross-validation. More importantly, the proposed method not only predicts which students will fail the course but also which students will achieve marginal results. This is an important advantage because marginal students are prone to failing. Indirect comparison between the RTV-SVM and categories (ii)-(iv) suggests that the proposed method can achieve superior OA, partly because of the greater complexity of these categories.

As a result, it can be seen that the proposed RTV-SVM can predict both at-risk and marginal students accurately, with a reduction of

Table 6
Performance comparison between proposed methods and other related methods.

Work	Aim	Target Group	Dataset	Method	Performance
(Lakkaraju et al., 2015)	Predicting at-risk students not finishing high school on time	U.S.; High school	Two schools; about 180000 students	Random forests Adaboost Logistic regression Support vector machine Decision tree	Accuracy: 0.8 Accuracy: 0.8 Accuracy: 0.77 Accuracy: 0.75 Accuracy: 0.72
(Hu et al., 2014)	Predicting at-risk students who fail the course	Taiwan; University	300 students (284 passed & 16 failed)	AdaBoost + C4.5 AdaBoost + CART	Accuracy: 0.972; Type I error: 0.007; Type II error: 0.049 Accuracy: 0.972; Type I error: 0.009; Type II error: 0.048
(Duarte et al., 2014)	Predicting at-risk students of dropping out	Portugal; University	293 students	Logistic regression	Accuracy: 0.909; Sensitivity: 0.933; Specificity: 0.855
(Gašević et al., 2016)	Predicting the student percent mark and thus at-risk student (mark less than 50%)	Australia; University	4134 students in 9 undergraduate courses	Multiple linear regression model	Overall R^2: 0.162
(Marbouti et al., 2016)	Predicting at-risk students who fail the course	U.S.; University	2973 students	Logistic regression Support vector machine Decision trees Multi-Layer perception Naïve based classifier K-Nearest Neighbor	Accuracy: 0.586 Accuracy: 0.724 Accuracy: 0.448 Accuracy: 0.483 Accuracy: 0.862 Accuracy: 0.345
(Jia & Maloney, 2015)	Predicting students of first-year course non-completions and second-year non-retentions	New Zealand; University	15833 students	Maximum likelihood probit model	Non-completion in first year Accuracy: 0.455 Non-retention in second year Accuracy: 0.414
(Alharbi et al., 2016)	Predicting students with poor honour	U.K.; University	984 students	Autoclassifier from IBM SPSS Modeler v.15	Accuracy: 0.67
(Kuzilek et al., 2015)	Predicting at-risk students who will not submit the next assessment	U.K.; University	32593 students	CART + Naïve based classifier K-Nearest + Neighbor	Accuracy: Varying from 0.476 to 0.934
(Proposed RTV-SVM)	Predicting at-risk students and marginal students who will receive fail grade or marginal grade in the course	U.K.; University	32593 students	RTV-SVM	Predicting at-risk students S_e: 92.1–94%; S_p: 92–93.6%; OA: 92.2–93.8% Predicting at-risk students and marginal students (137) S_e: 91.6–93.7%; S_p: 91–93.3%; OA: 91.3–93.5%

training vectors (and thus training time) while retaining the accuracy of the classifier.

6. Conclusion

This paper draws attention to the need to predict both the at-risk and marginal university students, and propose a reduced training vector-based support vector machine (RTV-SVM) for their accurate prediction. Performance evaluation showed the proposed classifier to achieve sensitivity of 92.1–94%, specificity of 92–93.6% and overall accuracy of 92.2–93.8% in predicting at-risk students, and sensitivity of 91.6–93.7%, specificity of 91–93.3% and overall accuracy of 91.3–93.5% in predicting marginal students. Moreover, it has been demonstrated herein that the RTV-SVM can reduce the number of training vectors, and thus the training time of the classifier by at least 60% while preserving the accuracy of the classifier. Accordingly, it is recommended that the RTV-SVM be adopted to reduce the training time when the size of the given dataset is large.

The major limitations of the proposed work are (i) there is room for the improvement of classification accuracy; (ii) the proposed RTV-SVM can be further tested to investigate it achieves outstanding performance in other learning analytics applications; and (iii) RTV-SVM has advantage on large dataset because it is trivial that many training vectors are selected as support vectors in small-scale dataset.

Aforementioned in section 2, besides predicting students who will fail the course, there are three other learning analytics applications, identifying students likely to defer their studies or fail to finish them on time, predicting which students will drop out, and identifying students who will graduate with a narrow pass, which could be the future directions to provide better support and experience in students' studies.

References

Alharbi, Z., Cornford, J., Dolder, L., & De La Iglesia, B. (2016). Using data mining techniques to predict students at risk of poor performance. *SAI Computing Conference,* 523–531.

Assessment Handbook for undergraduate and postgraduate module (pp. –). (2017). UK: The Open University.

Bertsekas, D. P. (1999). *Nonlinear programming.* Belmont: Athena Scientific.

Cemalcilar, Z., & Gökşen, F. (2014). Inequality in social capital: Social capital, social risk and drop-out in the Turkish education system. *British Journal of Sociology of Education, 35*(1), 94–114.

Chia, Y. Y., Lee, L. H., Shafiabady, N., & Isa, D. (2015). A load predictive energy management system for supercapacitor-battery hybrid energy storage system in solar application using the Support Vector Machine. *Applied Energy, 137*, 588–602.

Chui, K. T., Tsang, K. F., Chi, H. R., Ling, B. W. K., & Wu, C. K. (2016). An accurate ECG based transportation safety drowsiness detection scheme. *IEEE Transactions on Industrial Informatics, 12*(4), 1438–1452.

Chui, K. T., Tsang, K. F., Wu, C. K., Hung, F. H., Chi, H. R., Chung, H. S. H., et al. (2015). Cardiovascular diseases identification using electrocardiogram health identifier based on multiple criteria decision making. *Expert Systems with Applications, 42*(13), 5684–5695.

Dias, M. L. D., Neto, A. R. R., Dias, M. L. D., & Neto, A. R. R. (2017). Training soft margin support vector machines by simulated annealing: A dual approach. *Expert Systems with Applications, 87*, 157–169.

Duarte, R., Ramos-Pires, A., & Gonçalves, H. (2014). Identifying at-risk students in higher education. *Total Quality Management and Business Excellence, 25*(7–8), 944–952.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68–84.

Gibbs, D. B., & Culleiton, A. L. (2016). A project to increase educator cultural competence in mentoring at-risk nursing students. *Teaching and Learning in Nursing, 11*(3), 118–125.

Gotoh, J. Y., & Uryasev, S. (2017). Support vector machines based on convex risk functions and general norms. *Annals of Operations Research, 249*(1–2), 301–328.

Gu, B., & Sheng, V. S. (2017). A robust regularization path algorithm for ν-support vector classification. *IEEE Transactions on Neural Networks and Learning Systems, 28*(5), 1241–1248.

Hill, R. M., Taroslavsky, I., & Pettit, J. W. (2015). Enhancing depression screening to identify college students at risk for persistent depressive symptoms. *Journal of*

Affective Disorders, 174, 1–6.
Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machine. *IEEE Transactions on Neural Networks, 13*(2), 415–425.
Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior, 36*, 469–478.
Jia, P., & Maloney, T. (2015). Using predictive modeling to identify students at risk of poor university outcomes. *Higher Education, 70*(1), 127–149.
Kaur, R., Kaur, N., & Sood, S. K. (2017). Security in IoT network based on stochastic game net model. *International Journal of Network Management, 27*(4).
Kirar, J. S., Agrawal, R. K., Kirar, J. S., & Agrawal, R. K. (2017). Composite kernel support vector machine based performance enhancement of brain computer interface in conjunction with spatial filter. *Biomedical Signal Processing and Control, 33*, 151–160.
Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., Vaclavek, J., & Wolff, A. (2015). OU analyse: Analysing at-risk students at the open university. *Learning Analytics Review,* 1–16.
Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., et al. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1909–1918). .
Letseka, M., & Maile, S. (2008). *High university drop-out rates: A threat to South Africa's future.* Pretoria: Human Sciences Research Council.
Lindt, S. F., & Blair, C. (2017). Making a difference with at-risk students: The benefits of a mentoring program in middle school. *Middle School Journal, 48*(1), 34–39.
Liu, R., Yang, B., Zhang, X., Wang, S., & Chen, X. (2016). Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. *2016. Mechanical Systems and Signal Processing, 75*, 345–370.
Lytras, M. D., Raghavan, V., & Damiani, E. (2017). Big data and data analytics research: From metaphors to value space for collective wisdom in human decision making and smart machines. *International Journal on Semantic Web and Information Systems, 13*(1), 1–10.
Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education, 103*, 1–15.
Marin, L., Pawlowski, M. P., & Jara, A. (2015). Optimized ECC implementation for secure communication between heterogeneous IoT devices. *Sensors, 15*(9), 21478–21499.
Meng, J., Luo, G., & Gao, F. (2016). Lithium polymer battery state-of-charge estimation based on adaptive unscented Kalman filter and support vector machine. *IEEE Transactions on Power Electronics, 31*(3), 2226–2238.
Nandi, I., Srivastava, P. K., Shah, K., Nandi, I., Srivastava, P. K., & Shah, K. (2017). Floodplain mapping through support vector machine and optical/infrared images from landsat 8 OLI/TIRS Sensors: Case study from varanasi. *Water Resources Management, 31*(4), 1157–1171.
Ogunyemi, D. (2005). Factors that correlate with the US Medical Licensure Examination Step-2 scores in a diverse medical student population. *Journal of the National Medical Association, 97*(9), 1258.
Okujeni, A., van der Linden, S., Suess, S., & Hostert, P. (2017). Ensemble learning from synthetically mixed training data for quantifying urban land cover with support vector regression. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*(4), 1640–1650.
Pham, B. T., Bui, D. T., Prakash, I., Nguyen, L. H., & Dholakia, M. B. (2017). A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS. *Environmental Earth Sciences, 76*(10), 371.
Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Educational Technology & Society, 17*(4), 117–132.
Schöne, C., Tandler, S. S., & Stiensmeier-Pelster, J. (2015). Contingent self-esteem and vulnerability to depression: Academic contingent self-esteem predicts depressive symptoms in students. *Frontiers in Psychology,* 6.
Stergiou, C., Psannis, K. E., Kim, B. G., & Gupta, B. (2018). Secure integration of IoT and cloud computing. *Future Generation Computer Systems, 78*, 964–975.
Sun, M., Zhang, D., Wang, Z., Ren, J., & Jin, J. S. (2016). Monte Carlo convex hull model for classification of traditional Chinese paintings. *Neurocomputing, 171*, 788–797.
Taravat, A., Frate, F. D., Cornaro, C., & Vergari, S. (2015). Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images. *2015. IEEE Geoscience and Remote Sensing Letters, 12*(3), 666–670.
Toetenel, L., & Rienties, B. (2016). Analysing 157 learning designs using learning analytic approaches as a means to evaluate the impact of pedagogical decision making. *British Journal of Educational Technology, 47*(5), 981–992.
Wilson, A., Watson, C., Thompson, T. L., Drew, V., & Doyle, S. (2017). Learning analytics: Challenges and limitations. *Teaching in Higher Education,* 1–17.
Yao, Y., Zhai, J., Cao, Y., Ding, X., Liu, J., & Luo, Y. (2017). Data analytics enhanced component volatility model. *Expert Syst. Appl. 84*, 232–241.
Zheng, J., Pan, H., & Cheng, J. (2017). Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mechanical Systems and Signal Processing, 85*, 746–759.