Research Article

A Method of Early Warning for Course Learning Based on SMOTE and OCSVM

Shuyan Yu¹ and Zhe Wei²

¹Shaoxing University Yuanpei College, Shaoxing 312000, Zhejiang, China
²School of Computer Science, Civil Aviation Flight University of China, Guanghan 618307, China

Correspondence should be addressed to Shuyan Yu; shuyanyu1231@qq.com

Received 5 August 2022; Revised 2 September 2022; Accepted 6 September 2022; Published 28 September 2022

Copyright © 2022 Shuyan Yu and Zhe Wei. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the application of big data, artificial intelligence, and other related technologies in the field of education, using machine learning to carry out early warning for course learning has become an effective means to improve teaching quality. However, in the scene of early warning, the samples are significantly less than the ordinary samples, and the general clustering or classification methods are difficult to achieve good results. Therefore, this paper proposes an early warning for course learning method based on SMOTE and OCSVM. First, collect and preprocess students’ college entrance examination information and online course learning information data. Second, use SMOTE algorithm to expanding the samples. Then, the OCSVM model is designed, the Gaussian kernel function is used, and the Lagrange multiplier is used to solve the optimization problem for the optimization objective. The qualified student samples are selected for learning, and the classifier is trained, so as to classify the student data and realize the early warning of course learning. Select recall and F1_Score to evaluate the model, and comparative experiments are carried out. From the experiment, it is clear that in most cases, the method proposed in this paper is superior to the original sample and traditional methods in recall rate and F1_Score.

1. Introduction

Nowadays, big data and artificial intelligence have been widely used in the field of education. Meanwhile, with the popularization of these technologies in the field of education, teaching supervision and teaching evaluation have become new hot spots. In April this year, the Ministry of Education and other five departments issued several opinions on strengthening the teaching management of online open courses in general colleges and universities, pointing out that we should make full use of new generation information technologies such as big data and artificial intelligence to strengthen the monitoring of the learning process [1]. In his speech, at the special seminar on the high quality development of education in the new era jointly held by the Ministry of Education and other three departments in July 2022, Minister Huai Jinpeng mentioned that we should pay close attention to the reform of education evaluation, deepen the comprehensive reform in the field of education, and thoroughly implement the national education digitalization strategic action [2, 3]. Early warning for learning refers to analyzing students’ learning background, learning behavior, test scores, and other relevant data according to certain standards, sending prompt signals to teaching staff and students according to the analysis results, and providing targeted intervention opinions to students with problems [4]. Early warning for course learning is one of the effective means to supervise the teaching process and ultimately improve the teaching quality. Early warning of curriculum can help teachers get effective teaching feedback, help teachers remind and intervene students with curriculum learning difficulties as soon as possible, and provide basis for teaching management and decision-making. On the other hand, it can also help students to find potential learning problems as soon as possible, remind them to improve their learning methods and habits, and hope to complete the course successfully.
This paper proposes a method of early warning for course learning based on SMOTE and OCSVM, which can use students’ personal information and online learning data to predict students’ course learning. The effectiveness of this method is verified by comparative experiments with traditional methods.

2. Related Work

Taking “web of science core set” as the search source and “early warning for study” or “early warning for education” as the subject word, the documents from 2005 to 2022 were queried, with a total of 403 records, 6068 citations, 6584 citations, and 16.34 citations per article, as shown in Figure 1. Let "early warning for study" be the keyword in the database of CNKI, we found that there were about 318 relevant documents from 2005 to 2022. The number of documents issued annually is shown in Figure 2. On the whole, whether at home or abroad, the research on early warning for learning is on the rise.

After combing and analyzing the literature at home and abroad, we find that most of the literature starts from the field of education, takes learners as the center, and finds the problems in the learning process through the learning process data [5-8]. Data mining, machine learning, and deep learning technologies are the most widely used methods in early warning for learning [9-11]. However, the existing studies mostly elaborate the use of machine learning or deep learning to design learning early warning models from a macro perspective. There are few studies on how to apply the model to learning early warning scenarios [12]. In all articles, there are very few specific machine learning or deep learning algorithms mentioned. Some scholars have studied the application of SVM algorithm model in learning early warning [13], and some have used variational auto_encoder in learning early warning system [14]. Regarding the specific early warning for course learning, most existing research works are about the design and study at the theory level, while few are concerned about the practical design and technical implementation.

In addition, there are some problems in the early warning for course learning scenario, such as the small number of original samples and the imbalance of samples. Generally, the proportion of students who need early warning in all courses is not large, so the effect of ordinary clustering or classification technology is not very ideal. In
view of the small sample size in the early warning for course learning problem, the general random expansion method is likely to lose the characteristics of a small number of samples and affect the detection effect, while the expansion of SMOTE method can effectively maintain the proportion of a small number of samples. We can understand the early warning problem of course learning as two categories: need early warning and do not need early warning. After identifying any one category, the rest is another category. Therefore, it can be summed up as a kind of learning problem. So, we propose a method of early warning for course learning based on SMOTE and OCSVM, which combines and optimizes the two algorithms and applies them to early warning.

3. Principle of the Method

3.1. SMOTE Algorithm. SMOTE (synthetic minority over sampling technique) algorithm is an oversampling method proposed by Chawla et al. [15]. It is an improved scheme based on random oversampling algorithm [16]. The algorithm generates new samples through the relationship between samples; that is, the minority samples are obtained by random interpolation between the minority samples and their adjacent samples, so as to expand the samples [17]. SMOTE sampling can effectively alleviate the disadvantage of repeatedly adding the same samples in the random oversampling method, improve the imbalance of the original samples, reduce the over fitting of the model, and avoid the loss of features [18].

The simulation process of SMOTE algorithm samples the K-nearest neighbor (KNN) classification algorithm. First, for each sample in the minority sample, the distance from it to all samples in the minority sample is calculated, and K-nearest neighbors are obtained. Second, the ratio n is determined by the proportion between samples. Then, among the k-nearest neighbors, n samples of each minority sample are randomly selected. Finally, random linear interpolation is performed on these n samples to construct a new sample set as shown in Figure 3.

In the machine learning model, sufficient data can submit the accuracy of prediction. However, in the early warning for course learning, generally speaking, the number of students who need early warning is relatively small, so there is a risk of sample imbalance. If using unbalanced small samples to train the model, it will have a serious impact on the accuracy of the model, resulting in overfitting or underfitting and other problems [19]. Therefore, in the method of early warning for course learning of this paper, SMOTE algorithm is used to expand the training set.

The specific algorithm is as follows:

Let the number of samples of a minority class in the training set be Total, and then, SMOTE algorithm will synthesize N*Total new samples for this minority class sample. Where N is a positive integer. To prevent input errors, when N<1, the algorithm forces N = 1. Consider one of the few samples whose eigenvector is \(x_i, i \in \{1, \ldots, \text{Total}\}\):

1. First, find \(k\) nearest neighbors of sample \(x_i\) (e.g., Euclidean distance) from all the total samples of the minority sample, and record them as \(x_{i_{(\text{near})}} \in \{1, \ldots, \text{Total}\}\).
2. Second, a sample \(x_{i_{(\text{near})}}\) is randomly selected from the \(k\) nearest neighbors and regenerated into a random number \(r\) between 0 and 1, so as to synthesize a new sample \(x_{i_1}\):
   \[
   x_{i_1} = x_i + r (x_{i_{(\text{near})}} - x_i).
   \] (1)
3. Repeat step 2 for \(N\) times, so that \(n\) new samples can be synthesized: \(x_{\text{new}_i, \text{new} \in 1, \ldots, N}\).

By performing the above operations on the above total minority samples, the minority samples can be synthesized into \(N*\text{Total}\) new samples.

3.2. OCSVM Algorithm. SVM (support vector machine) is a kind of generalized linear classifier that classifies data according to supervised learning. It is widely used in pattern recognition, such as text classification, portrait recognition, and other issues [20]. However, if the number of samples in the training set is uneven, the classification interface of SVM method will tilt, and the final classification performance will decline [21]. However, in practical application scenarios, the problem of sample imbalance exists widely. In order to solve these problems, researchers proposed OCSVM.
OCSVM (one-class support vector machine) is a machine learning method proposed by Scholkopf et al. It is an extension of support vector machine (SVM) [22]. Unlike SVM, which focuses on two classifications, OCSVM only focuses on one classification, so it is particularly suitable for solving problems such as anomaly detection. Its basic idea is to map the target sample points to the corresponding feature space according to the corresponding kernel function, then construct a hyperplane between the data and the origin, maximize the distance between the target sample points and the origin, and finally return to a decision function to judge the category of samples [23, 24] as shown in Figure 4.

OCSVM model is described as follows. Set the sample set $S = \{x_i, i \in 1, \ldots, N\}$, map to the high-dimensional feature space through the kernel function $\phi$, and construct an optimal hyperplane in the feature space to maximize the distance between the target sample point and the coordinate origin. The coordinate origin is assumed to be the only abnormal sample, and the optimal hyperplane is as shown in the straight line in the figure, allowing a small number of samples between the coordinate origin and the interface. Weight of support vector (slope of hyperplane) $\omega$ and threshold (intercept of hyperplane) $\rho$. It is transformed into the following secondary planning problems:

$$
\min_{\omega, \xi, \rho} \frac{1}{2} \|\omega\|^2 + \frac{1}{vN} \sum_{i=1}^{N} \xi_i - \rho,
$$

$$
s.t. (\omega \cdot \phi(x_i)),$$

where $N$ is the size of the training set, $\xi$ is the relaxed sample variable, and $v$ is the regularization parameter, usually $v \in (0, 1)$, which is used to control the proportion of support vectors in the sample set. Introduce Gaussian kernel function:

$$
k(x_i, y_j) = \langle \phi(x_i), \phi(y_j) \rangle = \exp\left(-\frac{\|x_i - y_j\|^2}{2\sigma^2}\right).$$

Use Lagrange multiplier method to transform the above quadratic programming problem:

$$
\min_{\alpha} \sum_{i,j} a_i a_j k(x_i, x_j),
$$

$$
s.t. 0 \leq \alpha_i \leq \frac{1}{vN},$$

$$
\sum_{i} \alpha_i = 1.
$$

So weight $\omega = \sum a_i \phi(x_i)$ for either $a_i$, threshold $\rho = \sum_{i} a_i k(x_i, y_j)$. It can be concluded that the corresponding samples $0 \leq a_i \leq 1/vN$ are the relevant samples to determine the hyperplane, that is, the support vector. Therefore, the decision function is obtained:

$$
f(x) = \sum_{i=1}^{N} a_i k(x, x_i) - \rho.
$$

### 4. Early Warning Method for Course Learning Based on SMOTE and OCSVM

This paper proposes a method based on the combination of SMOTE and OCSVM. First, the sample is expanded by using SMOTE method, and then, the OCSVM model is designed to learn the sample data. Using the expanded training set, the classifier is trained, so as to realize the early warning mechanism of course learning.

#### 4.1. Data Set and Data Preprocessing

The experimental data set has 398 pieces of data related to the mobile application development course in 6 classes, includes the college entrance examination scores, volunteer, and learning general course platform of students. According to the research on the data, it is understood that there are features that are not beneficial to the model, these redundant features are deleted, and finally, seven numerical features are selected as training features. Data labels are set in the data set according to the final score. The final score $\geq 60$ is marked as 1, the final score $<60$ is marked as 0, and the students marked as failing are the students who need early warning.

Before starting the experiment, do some pretreatment on the data.

1. Denoise the data set and fill in the missing values. In this experiment, the missing values of "video viewing rate," "chapter learning rate," "average homework score," and "sign in rate" are set to 0, and other missing values are replaced with the average values of corresponding features.

2. Standardize the data set. In order to unify the data dimension, prevent gradient explosion or gradient dispersion, eliminate singular points, speed up convergence, eliminate the negative impact of noise data on the training model, and prevent over fitting, data standardization is generally carried out before data training. Data standardization includes maximum standardization, mean variance standardization, and quartile standardization. In addition, there are some normalization processing. In this paper, the sampling mean variance standardization method rescaled the features to zero mean and unit variance to standardize the data:

$$
x \leftarrow \frac{x - \mu}{\sigma}.
$$

$\sigma$ is variance, and $\mu$ is the mean value. After the above processing, these features have zero mean and unit variance.
4.2. Expanding the Sample by the SMOTE Algorithm. After preprocessing the above 398 pieces of data, it was found that tags marked with 0 accounted for 32.10%. From the perspective of a single class, the lowest proportion of 0 Tags is 11.50%, and the highest is 41.10%. There is a risk of sample imbalance. SMOTE algorithm is used for sample expansion. According to the SMOTE algorithm described above, the generated quantity is set: gen_num = 1000, nearest neighbor $k = 5$. The data set is divided into training set and test set in a ratio of 7:3. In the constructed training set, the type of failing the final grade accounts for 33.15% of all training samples.

4.3. Model Training by OCSVM Algorithm. For the new training samples generated by SMOTE above, the data set conforms to the unbalanced characteristics of samples. Because there are only two categories of data: 0 and 1, we use the data features of the category labeled 1 to train and get the hyperplane. The data that conform to these characteristics are judged as qualified, and the data that do not conform to these characteristics are judged as unqualified.

During training, we set kernel = "rbf"; that is, Gaussian distribution is adopted, and gamma correlation coefficient is set to 0.1. The training error score nu is set to 0.1 and 0.05, respectively, and the test is done. From the effect, nu = 0.05 is better.

5. Analysis of Experimental Results

The experimental equipment in this paper is a personal computer. The basic configuration is as follows: Intel (R) core (TM) i7-6700hq CPU @ 2.6 GHz 2.59ghz, 8 GB memory, 64 bit Windows10 operating system. Tensorflow is used for the programming.

In order to verify the effect of the experiment, this paper compares naive Bayes, OCSVM and the method in this paper. OCSVM adopts nu = 0.05.

There are too few abnormal samples in the early warning for course learning, which will lead to the problem of high accuracy but low recall rate in the test. Early warning learning problem requires identifying normal value samples and abnormal value samples, which can be considered as a binary classification problem in essence. F1_Score, as an index to measure the accuracy of binary classification model, takes into account the accuracy and recall of the model, and is widely used to evaluate the effect of machine learning model [25].

Use the confusion matrix to express the classification results, as shown in Table 1:

$$\text{PR} = \frac{TP}{TP + FP},$$

$$\text{AR} = \frac{TP + TN}{TP + FN + FP + EN}$$

$$\text{RR} = \frac{TP}{TP + FN}$$

$$\text{FS} = 2 \times \frac{\text{PR} \times \text{RR}}{\text{PR} + \text{RR}}$$

In the comparative experiment, the performances of several methods in the training set and the test set are shown in Table 2 and Table 3 respectively:

From the experimental results given in Tables 2 and 3, it can be seen that the method proposed in this paper is superior to the other two models that directly use the traditional algorithm in precision rate, accuracy rate, recall rate F1_Score, which verifies the effectiveness of the method.

6. Conclusions

As one of the important means of learning support, early warning for course learning has been paid more and more attention to by the educational community, especially with the construction of online open courses and the promotion of educational informatization, various platforms have accumulated a large amount of learning data, which provides data preparation for learning analysis. The course learning early warning method proposed in this paper collects the learning data on the learning pass, combines the students' personal college entrance examination information, adopts the combination of SMOTE and OCSVM algorithm, uses the Gaussian kernel function, and uses the Lagrange multiplier to solve the optimization problem of the optimization goal, which solves the problem of sample imbalance and improves the classification effect of the classifier. Through comparative experiments, in most cases, the method...
proposed in this paper is superior to the original sample and traditional methods in recall rate and F1 Score.

At present, the number of features in the method proposed in this paper is small. It only focuses on the college entrance examination information and students’ online learning data. In the future, we can collect information about students’ offline classes, such as attendance rate, frequency of answering questions, and even capture students’ facial expressions, movements and other data, and design more dimensional features to improve the early warning effect. With the progress of the Internet of things and information technology, we believe that learning early warning technology will be more accurate and widely used.

Data Availability

The data set used in this paper is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflicts of interest regarding this work.

Acknowledgments

This work was partially supported by the Second Batch of Collaborative Education Project of Industry University Cooperation under grant no. 202102281036, Scientific Project of CAFUC under grant nos. JG2022–06 and J2022–042, Sichuan Education Reform Project under grant nos. JG2021–521, Central University Education Reform Project under grant no. E2022078, and Sichuan Science and Technology Program under grant nos. 2022YFG0190 and 2022JDR0116.

References

[1] Ministry of education etc, “Opinions of the Ministry of education and other five departments on strengthening the teaching management of online open courses in general colleges and Universities,” 2022, http://www.gov.cn/zhengce/zhengce/2022-04/01/content_5682923.htm.

[2] Ministry of education, “The Ministry of education and other three departments jointly held a special seminar on the high-quality development of education in the new era,” 2022, http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/moe_1485/202207/7f20220723_648084.html.

[3] Y. Cao, “Research on risk prediction model construction and learning strategy of distance learning,” Region - Educational Research and Reviews, vol. 3, no. 4, pp. 10–12, 2021.

[4] W. Xiao, C. Ni, and R. Li, “Foreign research on learning early warning based on Data Mining: review and Prospect,” Distance education in China, vol. 02, pp. 70–78, 2018.

[5] L. Cassells, “The effectiveness of early identification of ‘at risk’ students in higher education institutions,” Assessment & Evaluation in Higher Education, vol. 43, no. 4, pp. 515–526, 2018.

[6] Y. H. Hu, C. L. Lo, and S. P. Shih, “Developing early warning systems to predict students’ online learning performance,” Computers in Human Behavior, vol. 36, pp. 469–478, 2014.

[7] H. Yang, J. Diao, T. Zhou, Z. Yao, X. Shi, and Z. Wang, “Research and implement of course early warning system based on teaching behaviour data,” Journal of Physics: Conference Series, vol. 1176, pp. 042052–052, 2019.

[8] D. Baneres, M. E. Rodriguez, and M. Serra, “An early feedback prediction system for learners at-risk within a first-year higher education course,” IEEE Transactions on Learning Technologies, vol. 12, no. 2, pp. 249–263, 2019.

[9] S. Lee and J. Y. Chung, “The machine learning-based dropout early warning system for improving the performance of dropout prediction,” Applied Sciences, vol. 9, no. 15, p. 3093, 2019.

[10] E. Howard, M. Meehan, and A. Farnell, “Contrasting prediction methods for early warning systems at undergraduate level,” The Internet and Higher Education, vol. 37, pp. 66–75, 2018.

[11] L. Wentworth and J. Nagaoka, “Early warning indicators in education: innovations, uses, and optimal conditions for effectiveness,” Teachers College Record, vol. 122, no. 14, pp. 1–22, 2020.

[12] M. Liz-Dominguez, M. Caeiro-Rodriguez, M. Llamas-Nistal, and F. A. Mikic-Fonte, “Systematic literature review of predictive analysis tools in higher education,” Applied Sciences, vol. 9, no. 24, p. 5569, 2019.

[13] Z. Mou, Y. Li, and D. Hu, “Design and implementation of learning early warning system based on learning behavior data in hybrid learning environment,” Distance education in China, vol. 36, no. 03, 2018.

[14] A. Abu Saa, M. Al-Emran, and K. Shaalan, “Factors affecting students’ performance in higher education: a systematic review of predictive data mining techniques,” Technology, Knowledge and Learning, vol. 24, no. 4, pp. 567–598, 2019.

[15] H. Lin, Y. Huang, D. Yang, K. Li, and J. Wei, “Application of improved multi-class Active learning algorithm in data processing,” in Proceedings of the 2021 International Conference on High Performance Big Data and Intelligent Systems (HPDBIS), pp. 116–124, Macau, China, December 2021.

[16] M. Lv, Y. Ren, and Y. Chen, “Research on imbalanced data based on SMOTE-AdaBoost algorithm,” in Proceedings of the2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), pp. 1165–1170, Xiamen, China, October 2019.

[17] X. Du, J. Yang, and J. L. Hung, “An integrated framework based on latent variational autoencoder for providing early warning of at-risk students,” IEEE Access, vol. 8, pp. 10110–10122, 2020.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, vol. 16, no. 1, pp. 321–357, 2002.

[19] X. Wu, J. Yang, and R. Nu, “A landslide susceptibility evaluation method combining SMOTE and convolution neural network,” Journal of Wuhan University (Natural Science Edition), vol. 45, no. 8, pp. 1223–1232, 2020.

[20] V. N. Vapnik, “An overview of statistical learning theory,” IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 988–999, 1999.

[21] G. E. A. P. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” SIGKDD Explor. News, vol. 6, no. 1, pp. 20–29, 2004.

[22] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the support of a high-
dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[23] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, Machinery Industry Press, Connecticut, NE, USA, 2019.

[24] L. A. Maglaras and J. Jiang, “A novel intrusion detection method based on OCSVM and K-means recursive clustering,” *ICST Transactions on Security and Safety*, vol. 2, no. 3, pp. e5–e10, 2015.

[25] R. Zhang, W. Chen, M. Hang, and L. Wu, “Abnormal flow detection of unbalanced samples based on variational self encoder in China Machine Press,” *Computer Science*, vol. 48, no. 7, pp. 62–69, 2021.

[26] D. Xiao, Q. Wang, M. Cai, Z. Zhu, and W. Zhao, “A3ID: an automatic and interpretable implicit interference detection method for smart home via knowledge graph,” *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2197–2211, 2020.