INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 16, Issue: 3, Month: June, Year: 2021
Article Number: 4153, https://doi.org/10.15837/ijccc.2021.3.4153

CCC Publications

An Approximation of Label Distribution-Based Ensemble Learning Method for Online Educational Prediction

Zhang Long, Shu Kai, Huang Keyu, Zhang Ruiqiu

Zhang Long
School of Design ,
South China University of Technology
382,Xiaoguwei Street, Panyu District, Guangzhou, Guangdong, CN510000, China
jim@belongto.cn

Shu Kai
University of Southern California,
3335 S Figueroa Street Los Angeles, CA, USA 90007,4244279094
kaishu@usc.edu

Huang Keyu
Zhejiang University of Finance and Economics,
Xueyuan Street, Xiashagaojiao District, Jianggan District, Hangzhou, Zhejiang, CN 310012
keyuhuang@belongto.cn

Zhang Ruiqiu*
School of Design ,
South China University of Technology
382,Xiaoguwei Street, Panyu District, Guangzhou, Guangdong, CN510000, China
*Corresponding author: anlepenult2019@126.com

Abstract

Online education becomes increasingly important since traditional learning is shocked heavily by COVID-19. To better develop personalized learning plans for students, it is necessary to build a model that can automatically evaluate students' performance in online education. For this purpose, in this study we propose an ensemble learning method named light gradient boosting channel attention network (LGBCAN), which is based on label distribution estimation. First, the light gradient boosting machine (LightGBM) is used to predict the performance in online learning tasks. Then The Channel Attention Network (CAN) model further improves the function of LightGBM by focusing on better results in the K-fold CrossEntropy of LightGBM. The results are converted into predicted classes through post-processing methods named approximation of label distribution to complete the classification task. The experiments are employed on two datasets, data science bowl (DSB) and answer correctness prediction (ACP). The experimental results in both datasets suggest that our model has better robustness and generalization ability.

Keywords: ensemble learning, light gradient boosting machine, channel attention network, CrossEntropy, label distribution approximation.

1 Introduction

COVID-19 has struck traditional education heavily, disturbing 94% students globally (99% of whom come from low and middle-income countries)[1]. Accompanied by continuous wars, traditional education has stagnated in the Middle East regions consequently as well [2]. Earthquakes and other similar natural disasters can also trigger a devastating blow [3] to school education. Besides, it is known that teachers play a crucial role in traditional education. However, limited energy has hindered teachers from taking good care of each student [4]. All the above suggests that the traditional face-to-face education model is fragile in some cases [5]. Therefore, online education is becoming increasingly prevalent, the development of which can guarantee that all the students can study normally [6]. With the Internet, information transparency in online education model can alleviate the discrepancy among students due to geographical and economic factors [7]. Individuals' time distribution can be more flexible in online education that complement time-related shortcomings in conventional class teaching[8]. Therefore, online education is attracting more and more supporters among students with personalized teaching programs. An online educational platform *Homework Help*, for instance, has provided aid for over 3.36 million primary and middle school students, as it revealed, and the number of students is growing rapidly[9]. Although the online education model is potential and prospective compared with traditional education methods, Matthew Effect will make a discrepancy between students when the online education model is applied. When Matthew Effect occurs, students with good self-regulation will obtain better grades while those who have poor self-regulation will obtain worse grades. Many education institutions and organizations operate investigations via questionnaires to determine how online education affected students [10-11]. But subjective factors and inadequate sample volume cannot be ignored in the form of questionnaires. Comparatively, when implementing such investigations in the form of big data, the hurdles mentioned above can be avoided [12].

The fast progress of online education enlarges the analyzable data, making it possible to apply machine learning to analyze students' performance in order to promote the development of online education and provide better educational service [13-19]. However, it is difficult to cope with several problems in machine learning such as overfitting caused by variable data distribution [20], making that a unitary machine learning algorithm is inferior in processing data without obvious regularity and performs unsatisfactorily in accuracy and anti-overfitting compared with ensemble learning methods [21]. Therefore, ensemble learning methods such as XGBoost [22] and LightGBM [23] have been employed extensively in prediction and classification tasks.

CrossEntropy (CE) loss function is usually used to calculate the probability of each class, and the class with the highest probability is selected as the prediction class. However, the model obtained by this method is prone to overfitting and difficult to achieve satisfactory results. In this study, we propose a Light Gradient Boosting Channel Attention Network (LGBCAN) model based on the approximation of label distribution (ALD) to improve the performance of the model. Experiments have been conducted in data science bowl (DSB) and answer correctness prediction (ACP) datasets, respectively.

The main contributions of this study are as follows.

1. We proposed a light gradient boosting channel attention network method, which can improve the generalization of the model with only increasing a few calculations. This network is divided into two parts: LightGBM model and a convolutional neural network with channel attention.

2. A post-processing approach named the approximation of label distribution has been applied to transform prediction results into class information to obtain enhanced adaptability in classification.

The rest of the study is organized as follows. Section 2 describes the related work. Section 3 presents the proposed method in detail. Section 4 carries out experiments and gives discussion about the experimental results, followed by the conclusion in Section 5.

2 Related Work

The methods applied in online education analysis can be divided into conventional methods, the methods based on conventional and unitary machine learning, and those based on ensemble learning.

2.1 Conventional Online Education Analysis

Progress in information technology has positively impacted the application of online education [13][24]. High involvement in these educational activities can direct students' thinking in the class and stimulate their enthusiasm for learning [25]. Benta et al.[26] demonstrated that students are more willing to complete assignments through the assistance of online education. Individualized characteristics of learners can be better understood in online learning research methods realized by fine-grained analysis. In this way can we customize individualized study content[27]. But the statistical techniques in online education have disadvantages of complicated calculation and subjective weight choice in evaluation[27].

2.2 Online Education Analysis Based on Conventional and Unitary Machine Learning

The report of Digital Technology and Management Center in 2015 pointed out that the amount of available digital education data is increasing[28]. Such an increase renders machine learning applications to conduct research on online education, which helps teachers have a better understanding of students' performance. With machine learning and mathematical statistics, Kotsiantis et al. [29] predicted students' scores by regression approaches after online education. Incorporating machine learning and data analysis in the studying management system, the model constructed by Villegas et al.[4] improved the learning efficacy of students in the online class. The model can perceive students' learning progress during interactions and analysis, contributing to creating an individualized online education model to satisfy most students' requirements. These approaches compensate for traditional face-to-face education and offer teachers aid to accomplish their teaching tasks.

2.3 Online Education Analysis Based on Ensemble Learning

At present, with the maturation of algorithms in machine learning and the development in hardware, ensemble methods are applied extensively in practical research. For instance, XGBoost or LightGBM is utilized in the classification [30][31], predictions [32][33] and authoritative competitions. According to research, ensemble models access much better performance than unitary models in large datasets[34]. Simultaneously, there is rapid development of ensemble learning technologies applied in online education. A boosted-like online learning enhancement (BOLE) ensemble method based on a heuristic modification to adapt diversity-based online boost (ADOB) was proposed by Barros et al. [20]. Xu et al.[35] developed a two-layer structure containing multiple base predictors and cascaded ensemble predictors, which can predict students' performance in the future. Besides, they proposed a data-driven method based on latent factor model and probability matrix decomposition that used to discover relations between courses in the curriculum, which can build an effective prediction base with high efficiency. These methods have shown excellence in improving the accuracy of online learning research. On this basis, we proposed LightGBM, a more advanced ensemble method that can obtain further improvement in the generalization ability of the model through the class information in the label.

3 The LGBCAN Model Based on Label Distribution Estimation

We propose a Light Gradient boosting Channel Attention Network (LGBCAN) based on the approximation of label distribution. The procedure of the proposed methods is shown in Fig. 1. The particular ideas and steps in each part will be introduced in the following sections. First, we eliminate the inconsistencies between the training and test sets by data pre-processing, and convert the textual information into computationally friendly data by encoding. Then the features of the processed data are extracted and fed into the LGBCAN model. LGBCAN is divided into two parts, LightGBM and Channel Attention Network (CAN). The LightGBM part obtains multiple computation results by K-fold cross-validation, which will be input into CAN. CAN model is utilized for further improving

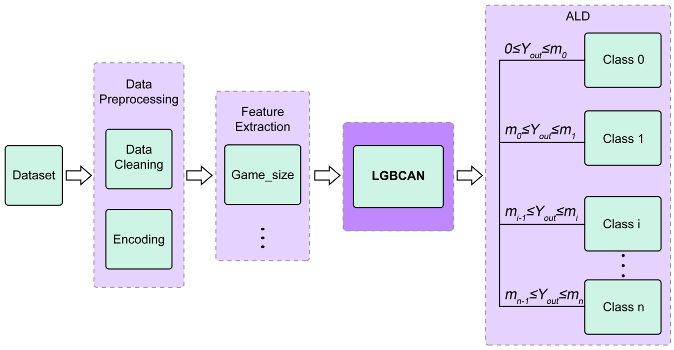

Figure 1: The Structure of Approximation of Label Distribution-Based Ensemble Learning Method for Online Educational Prediction

the generalization of the model. The methods can convert the prediction results to specific prediction classes during the ALD process.

3.1 LightGBM Prediction Model

LightGBM[23] is an ensemble learning algorithm that applies an improved histogram algorithm and a leaf-wise generation strategy to alleviate overfitting and avoid redundant features in computing. It is also endowed with predominant efficiency in running velocity and accuracy compared to conventional algorithms such as the gradient boosting decision tree (GBDT) [33]. Therefore, in this study, LightGBM is used to fulfill the prediction tasks in online education and serves as a basic function of the K-fold cross-validation (K=5) to mitigate overfitting and improve generalization.

In the prediction mission of time series, the functional relationship of discrepancy can be reflected by the loss function between prediction results and actual results, which is then utilized to evaluate the advantages and disadvantages of the model[36]. The loss function of LightGBM is computed by formula (1).

$$L = \sum_{1} l(\hat{y}_i, y_i) \tag{1}$$

In formula (1), $\hat{y}$ and y_i are respectively the prediction result and the real result of the *i-th* sample in a single branch of the decision tree, and l $(\hat{y}_i,\ y_i)$ is the loss function of a single branch of the decision tree.

However, due to the imbalanced data distribution in different classes in a dataset, the number of samples in the data centralized part is much higher than that of other samples. Consequently, the losses of the majority classes produces the majority of the total loss, so they are learned more completely, and the model has better accuracy in predicting the majority class. Nevertheless, when the model has a bias on major class samples, a severe scarcity emerges in learning minor classes. There are difficulties in achieving excellent results due to discrepancies in the data distribution. When such a model is applied in other datasets, over learning of the major samples unavoidably causes overfitting.

In this study, class weight is utilized in the loss function of the LightGBM model. The loss function can lower the weights of the major sample classes than those of the minor sample classes. And L1 regularization is introduced to reduce complexity and avoid overfitting. The related variation a_i is calculated in formula (2) and the rectified model loss function is depicted in formula (3).

$$a_i = floor(\frac{C}{n_i}) \tag{2}$$

$$L = \sum_{i} a_i \times l(\hat{y}_i, y_i) + \tau||\omega|| \tag{3}$$

In formula (2), n_i is the number of samples of the *i-th* class, a_i is the corresponding coefficient of class weight, and C is the total number of samples. In formula (3), $\tau||\omega||$is the regularization term of L1, ω is the automatically obtained parameter in the decision tree, τ is the coefficient of

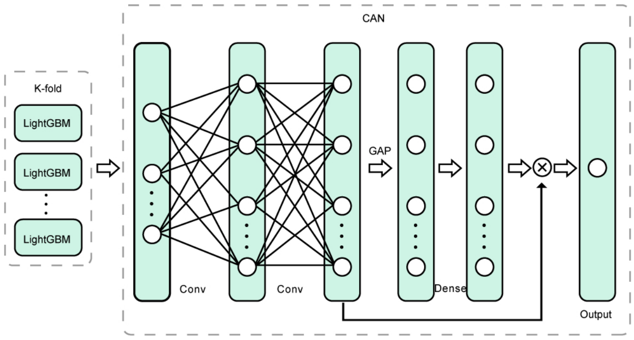

Figure 2: Structure of LGBCAN Model

the regularization term. Class weights are acquired by calculating the proportion of samples of each centralized class in all samples in the dataset; and the L1 regularization coefficient τ is obtained through Bayesian optimization. The L1 regularization coefficients searched by this method are less likely to be stuck in the local optimum.

3.2 The LGBCAN Model

Convolutional Neural Network (CNN) is a distinguished model in deep learning with extensive applications [37]. Applications of CNN are also hampered by overfitting in practice. To settle this problem, in this study the proposed LGBCAN consist of two convolutional layers of CNN and channel attention (CA). CA applied to optimized SENet'[38] learning outcome is introduced to the CNN, and the channel attention network (CAN) is utilized to enhance the training outcome of LightGBM. With negligible increase in time cost and memory, the generalization ability of the model is boosted. The detailed structure is shown in Fig. 2.

The attention mechanism strengthens significant characteristics and suppresses unrelated characteristics by weight analysis and other approaches. Five different prediction results of LightGBM obtained by K-fold cross-validation are employed as input of CAN. The part with the better regression results will adaptively obtain higher weights to improve the model's ability to predict students' performance. In contrast, the part irrelevant to the results is given lower weight.

3.3 Approximation of Label Distribution

The possibility of each class is solved out by the sigmoid function and the CE loss function in traditional classification methods. Despite extensive utilization, these methods are deficient in handling data overfitting. LightGBM first calculates the predicted output y_{out} whose value ranges between [0,1] by regression analysis, and comparison is implemented between y_{out} and the class thresholds. Lastly, the predicted class y_p of the model is calculated according to the thresholds. In a LightGBM model with a class number of *N*, the class y_p is computed by formula (4).

$$y_p \begin{cases} 0, y_{out} <= 1/N, \\ ..., \\ N, y_{out} <= 1, \end{cases} \tag{4}$$

In formula (4), y_{out} refers to the possibility of each sample class. This threshold setting method functions well when the number of samples in each class is equal. However, a large deviation will emerge when the number of samples in each class is unequal, which leads to a low model recall value and insufficient learning of minor sample classes. The possible overfitting in a dataset is neglected in this threshold setting method, so a label distribution approximation with self-adaptability is proposed

Table 1 Description of Class Information

Class	Description
0	Fail after over three attempts or more
1	Pass the assessment after three attempts or more
2	Pass the assessment after the second attempt
3	Pass the assessment after the first attempt

as formula (5).

$$m_i = \sum_{n=0}^{N} C_n/C \times 100(n \in [\ 0,\ N]) \tag{5}$$

$$y_p \begin{cases} 0, 0 \leq y_{out} \leq m_0, \\ ..., \\ N, m_{n-1} \leq y_{out} \leq m_n, \end{cases} \tag{6}$$

In formula (5), $C_0,\ C_1,\ \ldots,\ C_N$ is the number of samples in the 0, 1, ..., N-th class. The class's threshold is adjusted by self-adaptability in the number of samples to achieve more sufficient learning of the major sample classes. In formula (6), $m_0,\ m_1,\ \ldots,\ m_N$ is the threshold values of the 0, 1, ... , N-th class.

4 Experiments and Discussion

Some experiments are conducted to evaluate the proposed method. First, the datasets employed in experiments are introduced. Following are the parameters of the experiments and other detailed information. Finally, we discuss the experimental results and evaluates the proposed model.

4.1 Data Source and Analysis

To verify the performance of the proposed model, two data sets, DSB and ACP, are used in this study. Data preprocessing is realized by the ETL (Extract, Transform and Load).

4.1.1 The DSB Dataset

Booz Allen Hamilton and Kaggle jointly proposed The Data Science Bowl (DSB) competition [39]. Based on the given data, the number of children to pass the assessment can be classified.

According to the rules of the competition, the data in the dataset are divided into 4 classes, as Table 1 shows.

There are 11341042 records of children's game data in total in the training set, including 8294138 "valid" records which refer to the data that have been tried at least once. The training set and test set of the competitive data are children's actual performance on five evaluator tasks such as Bird Measurer, Cart Balancer, Cauldron Filler, Chest Sorter, and Mushroom Sorter, each corresponding to the children's understanding of a set of measuring-related skills.

4.2 The ACP Dataset

Provided by Riid Labs, the ACP dataset aims to help students living in difficult educational circumstances have equal access to educational resources when affected by factors such as COVID-19 [40]. It is a binary class competition, and its primary purpose is to predict students' performance in future interactions. The training data in the dataset includes 101,230,332 records in total belonging to 393,656 particular users. The final results about the users' answers are divided into two categories, namely, 0 referring to the wrong answer, and 1 referring to the correct answer. There is also a data distribution imbalance in this dataset, and the proportion of users who answered correctly reaches 64.45%, while users with the incorrect answers only accounted for 33.61%.

4.3 Metrics

It is far from adequate to evaluate a model only by accuracy as the metrics, which can cause a higher probability of misjudging a class with minor data as a class with the major data. In this study, F1 score, precision, accuracy and recall, Kappa coefficient are used as the criteria. For the sake of easy presentation, in the following part *Acc* means accuracy, *Pre* means precision, *Rec* means recall, *F1* means *F1_score* which is a harmonic average of accuracy and recall. *kappa* is the kappa coefficient, used to measure the consistency between the predicted class and the actual class.

These metrics are computed as formulae (7)-(12).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Pre = \frac{TP}{TP + FP} \tag{8}$$

$$Rec = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = \frac{2 \times Pre \times Rec}{Pre + Rec} \tag{10}$$

$$w_{i,j} = \frac{(i-j)^2}{(N-1)^2} \tag{11}$$

$$kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}} \tag{12}$$

In formulae (11) and (12), **O** and **E** are the actual class matrix and target prediction matrix, respectively, and w_i, j is calculated based on the difference between the actual and predicted class i with j, where N refers to the total number of classes.

4.4 Experiment Setting

We chose four benchmarks for comparison experiments. First are LightGBM (LGB) and XGBoost (XGB) methods that are widely used in DSB and ACP data sets. LGBALD and XGBALD are also used for the experiments in this study. Specifically, LGBALD is the classification method of LightGBM's regression prediction method combined with ALD method, and XGBALD is the classification method of XGBoost's regression prediction method combined with ALD method.

The experiments are conducted based on pytorch1.4 and python3.6 with Ubuntu 18.04 as the operating system. 30 epochs are trained in total. The learning rate of XGBoost and LightGBM methods is 0.01, and the initial learning rate of the proposed CNN model is 0.001. The decay of learning rate is observed from the 15th cycle to the 25th cycle with the decay rate 0.1, and the batch size is 128. For the fairness of the experiments, parameters are set as equal as possible in the model.

4.5 Experiment Results

4.5.1 Comparison Results in DSB Dataset

The metrics used in this experiment are *kappa* and the macro average values of *Pre*, *Rec*, and *F1*. The specific comparison results are shown in Fig.3 and Table 2. *kappa*, *Rec* and *F1* of the method proposed in this study are increased by 13.40%, 12.63%, 27.39% than those of LGBALD. At the same time, when comparing the proposed method with LGB and XGB, it can be observed that *kappa*, *Rec*, and *F1* are improved by 9.85%, 3.11% and 9.57% by the proposed model. The relatively low recall of each model is caused by a severe imbalance of samples. The improvement in the macro average *Rec* and *F1* of the model can be found compared with other models, which can alleviate the errors caused by the data imbalance to a certain extent. In summary, the model we proposed combines

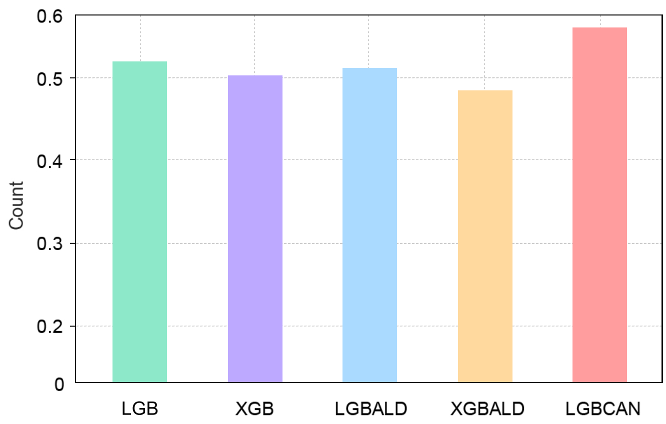

Figure 3: Kappa Coefficient in Difference Methods

the advantage of both the LGBALD and CNN. The LGBALD model has advantages in narrowing the difference between fitting points, and CNN focuses on the quick search for the best model in LightGBM. Therefore, as is shown in Table 2, the proposed method in most metrics outperforms the ensemble models in comparison.

4.5.2 Comparison Results in ACP Dataset

The performance of the proposed model can be manifested in the comparison results with the metrics of *Acc*, *Pre*,*Rec*, and *F1* which is shown in Table 3.

Acc is increased by 11.41% compared with that of LGBALD, and an obvious narrowed gap is observed between LGBCAN and LGB. *Pre* is increased by 5.56% compared with that of LGBALD and contributing to the obvious narrowed gap between LGBCAN and LGB. *Rec* is increased by 7.84% compared to that of LGB. *F1* is increased by 3.30% compared to that of LGB. It can be inferred that *Rec* and *F1* of the proposed method are increased by 7.84% and 3.30%, respectively, compared to LGB.

Besides, a conspicuous decrease in the discrepancy between LGBCAN and LGB is discovered in *Acc* and *Pre*, which are increased by 11.41%, 5.56% respectively, compared to LGBALD. The efficiency of the proposed model in eliminating overfitting and promoting generalization of the model are verified.

5 Conclusion

Online education has grown rapidly since the break of COVID-19, and it is necessary to build a model that can automatically evaluate students' performance in online education to develop personalized learning plans for students. In this study, we proposed an online education prediction method, and such a method predict the students' performance based on their historic performance. To improve the generalization ability of the model, an Approximation of Label Distribution-Based Ensemble Learning Method called LGBCAN is proposed. Experiments are conducted in two online learning related datasets: data science bowl (DSB) and answer correction process (ACP). The experimental results in both datasets show that the proposed LGBCAN model has better generalization ability in predicting students' performance in online education and its comprehensive performance is better.

Table 2 Comparison Results of Different Models

Model	Class	*Pre*	*Rec*	*F1*
LGB	class 0	0.6367	0.6403	0.6385
	class 1	0.3722	0.1262	0.1872
	class 2	0.5113	0.0034	0.0066
	class 3	0.6393	0.9921	0.7548
	macro-average	**0.5370**	0.4178	0.3907
XGB	class 0	0.6319	0.5994	0.6152
	class 1	0.3607	0.1074	0.1662
	class 2	0.2012	0.0016	0.0033
	class 3	0.6309	0.9142	0.7465
	macro-average	0.4559	0.4107	0.3888
LGBALD	class 0	0.8175	0.2046	0.3272
	class 1	0.2138	0.3259	0.2582
	class 2	0.1633	0.6348	0.2598
	class 3	0.7948	0.3647	0.5101
	macro-average	0.4973	0.3825	0.3363
XGBALD	class 0	0.8095	0.1779	0.2917
	class 1	0.2127	0.3531	0.2654
	class 2	0.1588	0.6449	0.2548
	class 3	0.7814	0.3061	0.4398
	macro-average	0.4906	0.3705	0.3130
LGBCAN (ours)	class 0	0.6045	0.6441	0.6237
	class 1	0.2305	0.2316	0.2312
	class 2	0.1499	0.1424	0.1463
	class 3	0.7184	0.7055	0.7119
	macro-average	0.4259	**0.4308**	0.4281

Table 3 Comparison Results of Different Models in ACP Dataset

Model	*Acc*	*Pre*	*Rec*	*F1*
LGB	**0.6833**	**0.6437**	0.5903	0.6158
XGB	0.6771	0.6245	0.5863	0.6047
LGBALD	0.5784	0.6373	0.5714	0.6026
XGBALD	0.5676	0.6323	0.5696	0.5993
LGBCAN	0.6814	0.6356	**0.636**6	**0.6361**

Author Contributions

The authors contributed equally to this work.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this study.

References

[1] https://www.un.org/development/desa/dspd/wpcontent/uploads/sites/22/2020/08/sg_policy_brief_covid19_and_education_august_2020.pdf

[2] K.D. Rajab, The Effectiveness and Potential of E-Learning in War Zones: An Empirical Comparison of Face-to-Face and Online Education in Saudi Arabia, in IEEE Access, vol. 6, pp. 6783-6794, 2018

[3] N. Todorova, N. Bjorn-Andersen, University learning in times of crisis: The role of IT. Accounting Education, vol. 20, no. 6, pp. 597–599. Dec 2011.

[4] W. Villegas-Ch, M. Román-Cañizares, and X. Palacios-Pacheco. Improvement of an online education model with the integration of machine learning and data analysis in an LMS, Applied Sciences, vol. 10, issue 15, pp. 1-18. August 2020.

[5] T. Chen, L. Peng, X. Yin, J. Rong, J. Yang, and G. Cong, Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic, Healthcare, vol. 8, no. 3, pp. 200, Jul 2020.

[6] D. Benta, G. Bologa, I. Dzitac. E-learning Platforms in Higher Education. Case Study, Procedia Computer Science, vol. 31, pp. 1170-1176, May. 2014.

[7] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn and N. Radi, Machine learning approaches to predict learning outcomes in Massive open online courses, 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 713-720.

[8] S. Houlden and G. Veletsianos. A post humanist critique of flexible online learning and its 'anytime anyplace' claims. British Journal of Educational Technology. 2019.

[9] H. Xu. https://new.qq.com/rain/a/20201018A028AO00, Oct. 2018.

[10] C. Xiao and Y. Li, Analysis on the Influence of the Epidemic on the Education in China, 2020 International Conference on Big Data and Informatization Education (ICBDIE), Zhangjiajie, China, pp. 143-147, 2020.

[11] B. McCarthy, L. Li, Tiu, M., Atienza, S. (2013). PBS KIDS mathematics transmedia suites in preschool homes. In Proceedings of the 12th International Conference on Interaction Design and Children (pp. 128–136). ACM.

[12] B. Thorns, E. Eryilmaz. "Introducing a twitter discussion board to support learning in online and blended learning environments", Education and Information Technologies, Vol. 20, No. 2, pp. 265-283. Jun. 2015.

[13] D. Gašević, C. Rose, G. Siemens, A. Wolff, and Z. Zdrahal, "Learning Analytics and Machine Learning," Proc. Fourth Int. Conf. Learn. Anal. Knowl. LAK, 14, pp. 287–288, 2014.

[14] J. Levy, D. Mussack, M Brunner, U Keller and P Cardoso-Leite, A Fischbach. Contrasting Classical and Machine Learning Approaches in the Estimation of Value-Added Scores in Large-Scale Educational Data. rontiers in Psychology, Aug 2020 2190.

[15] H. Vartiainen, M. Tedre and T Valtonen. Learning machine learning with very young children: Who is teaching whom?, International Journal of Child-Computer Interaction. Vol.9, No. 25, Sep. 2020.

[16] E. Dragan. "Interactive educational game using machine learning", In Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts pp. 272-275. June. 2020.

[17] Sano, Mina. Statistical Analysis of Elements of Movement in Musical Expression in Early Childhood Using 3D Motion Capture and Evaluation of Musical Development Degrees through Machine Learning, World Journal of Education, Vol. 8 No. 3, pp. 118-130, 2018.

[18] J. Hodges, S. Mohan. "Machine Learning in Gifted Education: A Demonstration Using Neural Networks", Gifted Child Quarterly, Vol. 63, No. 4, pp. 243-252. Sep. 2019.

[19] C. K Blackwell, A. R. Lauricella, E Wartella. "Factors influencing digital technology use in early childhood education", Computers and Education, Vol. 7, No. 7, pp. 82-90, Aug. 2014.

[20] R. S. M. d. Barros, S. Garrido T. de Carvalho Santos and P. M. Gonçalves Júnior, A Boosting-like Online Learning Ensemble, 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 1871-1878.

[21] H. Chen, S. Lundberg, S. Lee. Checkpoint Ensembles: Ensemble Methods from a Single Training Process, https://arxiv.org/abs/1710.03282, 2017.

[22] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system, In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp 785–794. Aug. 2016.

[23] G. Ke, Q. Meng, T Finley, T Wang, W Chen and W Ma. Lightgbm: A highly efficient gradient boosting decision tree, In: Advances in Neural Information Processing Systems. pp. 3149–3157. Dec. 2017.

[24] J. Qiu et al., Modeling and Predicting Learning Behavior in MOOCs, Proc. Ninth ACM Int. Conf. Web Search Data Min., pp. 93–102, 2016.

[25] P. Moreno-Ger, D. Burgos, I. Martínez-Ortiz, J. L. Sierra and B. Fernández-Manjón. Educational game design for online education. Computers in Human Behavior, Vol. 24, No. 6, pp. 2530–2540. Sep. 2008.

[26] D. Benta, G. Bologa, S. Dzitac, I. Dzitac, University Level Learning and Teaching via E-Learning Platforms, vol. 55, pp. 1366-1373, 2015.

[27] J. Wong, M. Baars, D. Davis, Tim Van Der Zee, Geert-Jan Houben, Fred Paas. Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review, International Journal of Human–Computer Interaction, vol. 35, issue 4-5, 356-373, 2019.

[28] V. Gamper and S. Nothelfer. The Future of Education Trend Report", Center for Digital Technology and Management. Jun. 2015.

[29] S. B. Kotsiantis. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades, Artificial Intelligence Review, Vol. 37, No. 4, pp.331-344, Apr. 2012.

[30] D. Wang Y. Zhang, and Y. Zhao. LightGBM: an effective miRNA classification method in breast cancer patients, Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics. pp. 7-11, Oct. 2017.

[31] Rong, E. Fonseca, D. Bogdanov, O. Slizovskaia, E. Gomez and Serra. Acoustic scene classification by fusing LightGBM and VGG-net multichannel predictions, roc. IEEE AASP Challenge Detection Classification Acoust. Scenes Events. pp.1-4, Nov. 2017.

[32] J Zhou, G Wang S Yang, J Liu, W Xu, Z Wang and J Ye. Automatic sleep stage classification with single channel EEG signal based on Two-Layer stacked ensemble model, IEEE Access, Vol. 8, pp. 57283-57297, 2020.

[33] Y Ju, G Sun, Q Chen, M Zhang, H Zhu and M U Rehman. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting, IEEE Access, vol.7: 28309-28318. Feb. 2019.

[34] E. PHUA, N. K. BATCHA. Comparative Analysis of Ensemble Algorithms' Prediction Accracies in Education Data Mining. JCR. Vol.7, Issue 3, pp. 37-40. July 2020.

[35] J. Xu, K. H. Moon and M. van der Schaar, A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs, in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 5, pp. 742-753, Aug. 2017.

[36] Z. Shi and M. Han, Support Vector Echo-State Machine for Chaotic Time-Series Prediction, in IEEE Transactions on Neural Networks, vol. 18, no. 2, pp. 359-372, March 2007.

[37] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, in Proceedings of the IEEE on, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[38] Jie Hu, Li Shen, Gang Sun. Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132-7141, 2018.

[39] https://datasciencebowl.com/

[40] https://www.kaggle.com/c/riiid-assessment-answer-prediction/data

Cite this paper as:
Zhang Long, Shu Kai, Huang Keyu, Zhang Ruiqiu, An Approximation of Label Distribution-Based Ensemble Learning Method for Online Educational Prediction, *International Journal of Computers Communications & Control*, 16(3), 4153, 2021.
https://doi.org/10.15837/ijccc.2021.3.4153