Title
SPOC learner’s final grade prediction based on a novel sampling batch normalization embedded deep neural network method

Journal
Multimedia Tools and Applications, 82(7)

ISSN
1380-7501

Authors
Liang, Zhuonan
Liu, Ziheng
Shi, Huaze
et al.

Publication Date
2023-03-01

DOI
10.1007/s11042-022-13628-y

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-13628-y

1222: INTELLIGENT MULTIMEDIA DATA ANALYTICS AND COMPUTING

SPOC learner's final grade prediction based on a novel sampling batch normalization embedded deep neural network method

Zhuonan Liang[1] · Ziheng Liu[1] · Huaze Shi[1] · Yunlong Chen[1] · Yanbing Cai[2] · Hong Hong[3] · Yating Liang[2] · Yafan Feng[1] · Yuqing Yang[4,5] · Jing Zhang[6] · Peng Fu[2]

Received: 30 January 2021 / Revised: 30 June 2022 / Accepted: 1 August 2022

Abstract
Recent years have witnessed the rapid growth of Small Private Online Courses (SPOC) which is able to highly customized and personalized to adapt variable educational requests, in which machine learning techniques are explored to summarize and predict the learners' performance, mostly focus on the final grade. However, the problem is that the final grade of learners on SPOC is generally seriously imbalance which handicaps the training of prediction model. To solve this problem, a sampling batch normalization embedded deep neural network (SBNEDNN) method is developed in this paper. First, a combined indicator is defined to measure the distribution of the data, then a rule is established to guide the sampling process. Second, the batch normalization (BN) modified layers are embedded into full connected neural network to solve the data imbalanced problem. Experimental results with other three deep learning methods demonstrate the superiority of the proposed method.

Keywords Grade prediction · Class balance · SPOC · Deep neural network · Batch normalization

1 Introduction

With the continuous development of the online education and the knowledge sharing projects, online sharing courses comes into public views [1, 2, 4]. Small Private Online Courses (SPOC) is one of the most popular solutions of implementing online sharing courses. SPOC benefits the educational source sharing without the constraints of time and location. As an online education method without face-to-face, tracking the learning situation of students play a key role in SPOC, guaranteeing the education effect. Predicting student final grade of courses is a

✉ Peng Fu
fupeng@njust.edu.cn

Extended author information available on the last page of the article

straightly approaching for investigating the teaching effectiveness. The advanced research made great contributions to precisely foresee the performance of student in the final exam. I. C. Juanatas et al. [8] presented the licensure examination performance prediction based on their academic grades using Logistic Regression. That provides a binary result liked PASS or FAIL, which is insufficient for further analysis [11]. For detailed prediction, the support vector machine (SVM) is applied. In specialized dataset, SVM shows outstanding performance compared to current methods [9]. While it is sensitive to the distribution of training datasets, lacking the practical generality. For traditional educational data under the massive scale, Y. Yang et al. [10, 14] proposed an improved random forest method to comprehensively predict the grade of students in the final exams. It is effective in prediction task for less students, but difficult to handle massive educational data in SPOC. The deep learning methods are applied in the educational system with large amount students [18]. A Bayesian deep learning model with variants was proposed for grade prediction under a course-specific framework [6]. While it is only applied to the offline or the traditional school education. The investigating features of online education are different from offline but providing an approaching to model the student behavior [12]. T. Yang et al. [13] proposed a model predicting the performance of students base on online video click stream events. While there are some obstacles in some cases of the SPOC grade prediction. The special circumstance on SPOC causes the imbalanced final grades of student which concentrate on upper score between 90 and full mark. This sort of long tail distribution causes derivation compared with the standard Gaussian distribution expected in the neural network, resulting in the negative effect on predication [3, 5, 15–17]. To address this problem, we propose a sampling batch normalization embedded deep neural network (SBNEDNN). In proposed model, we reconstructed an indictor to measure the distribution of the dataset, guiding the following sampling operation. Moreover, we embedded the batch normalization (BN) modified layers into the multilayer perceptions, furtherly handling the training difficulties caused by imbalance educational data [7]. Our contributions can be summarized as: 1) The paper proposes a data distribution indicator Max_score (MS) to measure the imbalance situation of education data. MS comprehensively considers the skewness index and kurtosis index of data. It is the guideline in the data processing procedure. 2) An improved imbalance data processing method is proposed in this paper. This method does not simply transform data into the uniform distribution. It reconstructs the resampling method to fit the hypothetical distribution in the following neural network design under the guideline of MS. 3) Experiments results with the widely used deep learning methods prove the superiority of our proposed method.

2 Sampling-BN embedded deep neural network method

In this part, we will describe the proposed model SBNEDNN. It includes two main components: data processing and neural predicting. To encounter the influence caused by distribution, a modified data process is designed, including diagnosing and shifting. For the diagnosing, a reconstructed indicator of data is applied to evaluate the distribution, integrating skewness and kurtosis rules. Then the adjustment of data using sampling is operated guided by the indicator, acting as the shifting module. After the processed, the data will be trained by the modified batch-normalized fully connected networks.

Data processing In the SBNEDNN method, we firstly evaluate the distribution of the dataset. The histogram presents qualitative analysis of the dataset distribution. However, it is not convincing to provide the quantitative analysis of the distribution, such that a define evaluation index is needed for guidance of the following sampling operation. Skewness, kurtosis, and standard deviation each measures the characteristics of data distribution. We constructed a quantitative indicator integrates three characteristics indicators above, diagnosing the data distribution and guiding the sampling. Specifically, the corresponding formulas are generated from the standard Gaussian distribution.

The skewness S can indicate the extent that a given distribution varies from a normal distribution, which can be written as Eq. (1):

$$S = \frac{1}{n}\sum\left[\left(\frac{X_i - \mu}{\sigma}\right)^3\right], \tag{1}$$

where S denotes the skewness, X_i represents the i_{th} student's score, μ represents the mean value of the final grade of all students and σ represents the standard variance of the scores of all students, n represents the number of students.

The skewness indicates the numerical imbalanced characteristics of data distribution. In practice, when $S < 0$, the probability distribution graph is biased to the left, on the contrast, when the $S > 0$, it is biased to the right. $S = 0$ is a good symbol that the data is relatively evenly distributed on both sides of the average value. Another measure in our indicator is kurtosis K, which indicates the steepness of the probability distribution of a random variable. The measure can be written as Eq. (2):

$$K = \frac{1}{n}\sum\left[\left(\frac{X_i - \mu}{\sigma}\right)^4\right] - 3, \tag{2}$$

where K denotes the kurtosis, μ represents the mean value of the scores of all students, σ represents the standard variance of the scores of all students, n represents the numbers of students. Specially, it assumes that the K value of the data being standard Gaussian distribution equals zero, which is the result of −3. When the kurtosis $K > 0$, it means that the data distribution is sharper than the normal distribution, on the other hand, $K < 0$ means that it is squat compared to the normal distribution. By combining S and K, σ, the developed indicator $Max\ \_\ score$ can be written as Eq. (3):

$$MS = \frac{\max(abs(S,\ K))}{\sigma}, \tag{3}$$

where $max(\cdot)$ means selecting the largest data value, $abs(\cdot)$ represents calculating the absolute value, S represents the skewness, K represents the kurtosis, σ represents the standard variance of MS. The proposed indicator presents how the data converge to Gaussian distribution considering both skewness and kurtosis. Moreover, we introduced the hypothesis testing to evaluate the data distribution via measuring the MS. Z-test formula is opted as the criteria in this part. The formula can be presented as:

$$Z = \frac{MS}{\sigma}, \tag{4}$$

where σ presents the standard value of MS. Z is the testing statistical value in the hypothesis test. Z reflects the difference between testing data distribution and Gaussian distribution, considering skewness and kurtosis through selecting the most distant indicator. To exemplify,

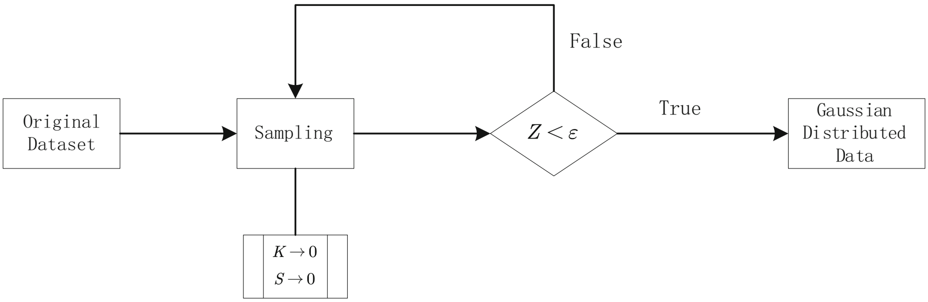

Fig. 1 The logistic of data preprocess. *K* presents the kurtosis and *S* presents the skewness. Z presents the evaluation value of *Max* _ *score*

the skewness and kurtosis of one specified dataset are 0.194 and 0.373, and σ in Z is obviously a value less than the standard values of skewness and kurtosis, representing as M which is 0.36 in this case. Z is expected ranging from 0 to 1.96 as statistical standard Z-test threshold value ε, while it is 1.036 in this example. According to the hypothesis testing result, it can be supposed that the data distribution obeys the standard Gaussian distribution.

When we obtain the S and K, *Max* _ *score* of the dataset of scores, we take S and K to zero as the direction of sampling and stop sampling until the value of Z is less than ε (usually set as 1.96). For better illustration, a flow chart is exhibited in Fig. 1. Under the guidance of the *Max* _ *score*, the classes with more samples are first processed by an under-sampling approach, the classes with less samples are first processed by using an over-sampling approach, which helps to balance class distribution by replicating data of minority sample. Figure 2 shows these two methods detailly.

Once the data is processed by above processes, the fully connected neural network with BN modification is applied in the subsequent prediction, the three steps of the proposed

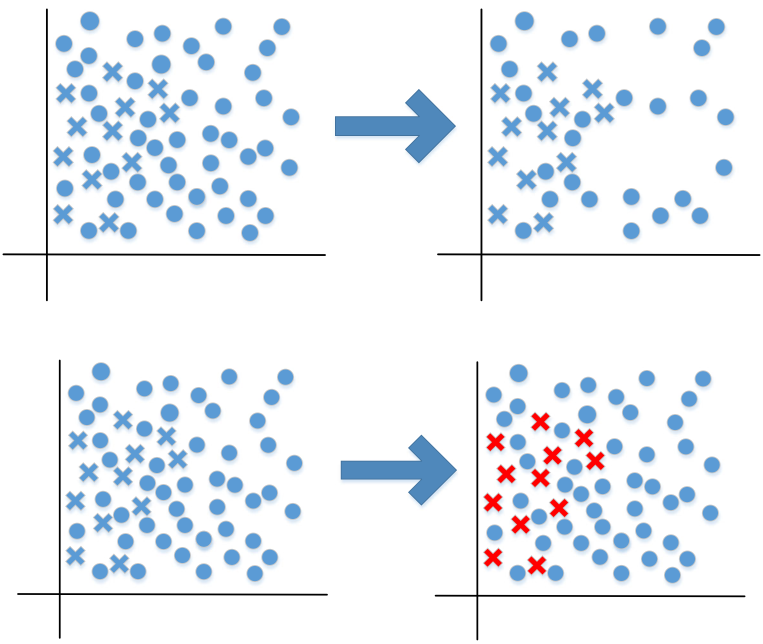

Fig. 2 The changes of data distribution from original to well-balanced after under sampling and over sampling

Table 1 The implementation process of the SBNEDNN Method

```
SBNEDNN Method
Step1: Data evaluation and preprocess.
do
{
        S = 1/n ∑[((X_i − μ)/σ)^3];
        K = 1/n ∑[((X_i − μ)/σ)^4] − 3;
        Max_score = max(abs(S, K))/σ;
        Z = Max_socre/σ;
        Procedure sampling
        {
                if (S < 0){
                        oversampling lower classes;
                }else{
                        oversampling upper classes;
                }
                if (K > 0){
                        undersampling medium classes;
                }else{
                        oversampling medium classes;
                }
        }
} while (Z > ε)

Step2: Score prediction
Split the training dataset into mini batches;
Loop {
        Loop (3){
                fully connected layer (mini batch);
                batch normalization layer (y);
        }
}
```

$$S=\frac{1}{n}\sum\left[\left(\frac{X_i-\mu}{\sigma}\right)^3\right];$$

$$K=\frac{1}{n}\sum\left[\left(\frac{X_i-\mu}{\sigma}\right)^4\right]-3;$$

$$Max\_score=\frac{\max\left(abs\left(S,\ K\right)\right)}{\sigma};$$

$$Z=\frac{Max\_socre}{\sigma};$$

SBNEDNN model can be summarized as Table 1. For a better illustration, the pre-processing procedure is represented by Fig. 3a, while the deep learning model prediction is displayed on the Fig. 3b and c.

3 Experiments

3.1 Dataset

We conduct the experiments on two SPOC dataset, Medic 2020 and Statistics 2020. Medic 2020 is a real-world dataset with 40,494 students from SPOC for medical graduate education on Chaoxing platform[1]. Statistics 2020 is a dataset with the desensitization learning record for the

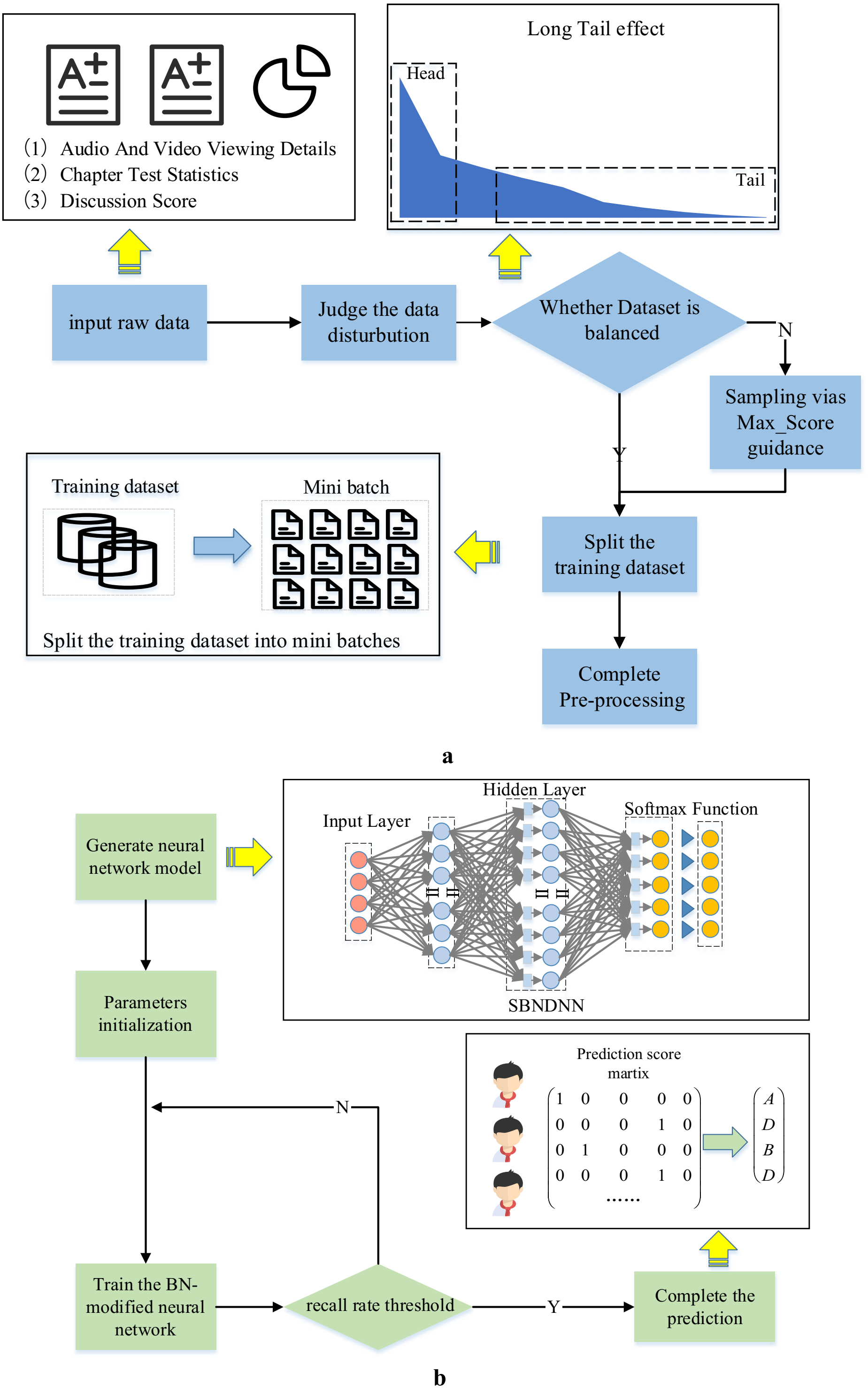

Fig. 3 **a** The pre-processing procedure of SBNEDNN Method **b** The model training procedure of SBNEDNN Method **c** The training network of SBNEDNN

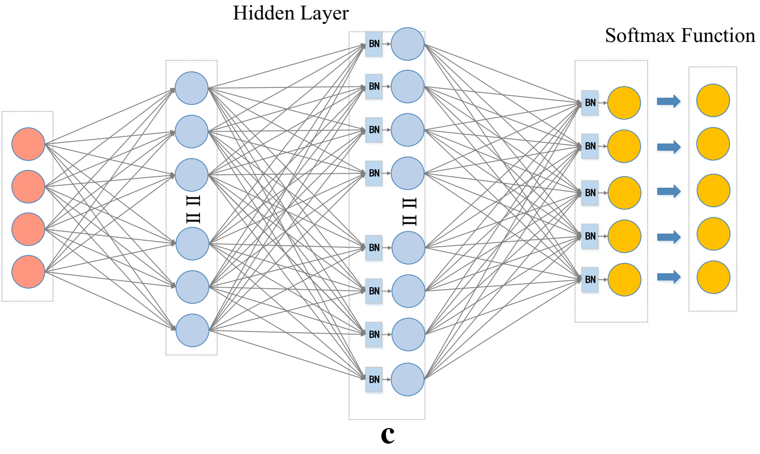

Fig. 3 continued.

year 2020 on the "Medical Statistics Program of the National Association of Medical College Graduate Schools SPOC Platform", including 5254 students. It tracks the learner's performance of exercises and tests during their studies on the online learning platform. The utilized SPOC datasets contain various features about students, including student background, learning behavior record, and test performance. To focus the learning-related features, we simplify the dataset, filtering out the background-related features, like age, hometown, and gender. We focus on the features reflecting the learning behavior of student on the SPOC platform. The filtered dataset has sixty-nine features, which can be classified into three categories, "Audio and Video", "Chapter Test Record", "Discussion Score". The detailed description of these features is presented in Table 2. In the SPOC datasets, some records of the features are missing in some specified students. The records of these student are dropped in the data preprocess. In the datasets, there are six final grades of student according to the score of final exam as Eq. 5. The percentage of final grades in two datasets is shown as Fig. 4. The data distribution of Medic 2020 is imbalanced as shown. The majority grade of Medic 2020 is L6, while the L4 takes the least. By contrast, the percentage of every grade on Statistics 2020 is fair comparing with the Medic 2020.

$$level = \begin{cases} L1 & 0 \leq \text{score} \leq 70 \\ L2 & 70 \leq \text{score} \leq 80 \\ L3 & 80 \leq \text{score} \leq 90 \\ L4 & 90 \leq \text{score} \leq 93 \\ L5 & 93 \leq \text{score} \leq 95 \\ L6 & 95 \leq \text{score} \leq 100 \end{cases} . \tag{5}$$

Table 2 Features Description

Three main feature classes	Description
Audio and Video (60 features)	The score of behavior in sixty lesson sections.
Chapter Test Statistics (8 features)	The score of each chapter test.
Discussion Score (1 features)	The score in the discussion activities.

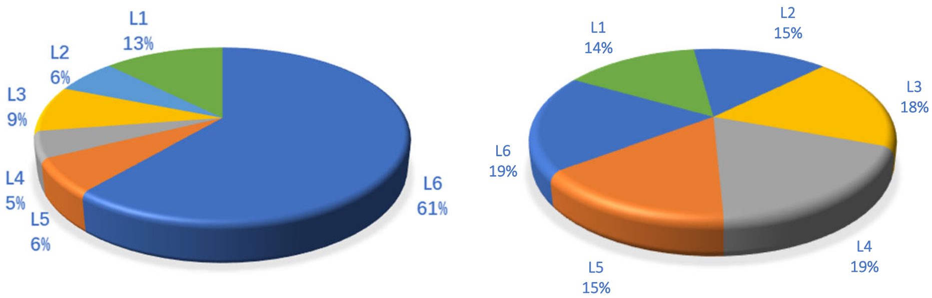

Fig. 4 The number of samples in each class with the experimental dataset. Left: Medic 2020 Right: Statistics 2020

3.2 Parameter setting and evaluation metrics

To initiate the model, we opt the Xavier initialization and generate the parameter from the gauss distribution. The average value and variance are calculated from the training dataset. The mini-batch size is 64. Adam algorithm is applied for optimization. The activation function is Sigmoid. The train-test split ratio is 70 versing 30. To measure the performance of different structures, we selected the prediction accuracy to evaluate whether model can predict the final grade of students.

We implement the all experiments with PyTorch via Python and conduct them on a Linux server with two 2.0GHz Intel Xeon E5–2683 CPUs and two GeForce 1080Ti GPUs.

3.3 Experimental results

3.3.1 Baseline experiments comparison

To show the superiority of the SBNEDNN method, three widely used deep learning methods, including CNN, RNN, LSTM, are applied with the pre-balanced dataset. The prediction accuracy of these methods is shown in Tables 3 and 4, the best performance of prediction task is in bold in both tables. The SBNEDNN method achieves the highest prediction accuracy compared to the other three methods in overall. Moreover, we can get further interesting information if taking educational practice into consideration. As Tables 3 and 4 shown, the proposed model shows the superior performance among the medium three classes. According to the educational experience, the students of the medium classes ranging 70 to 90 should be specially focused for promoting them or preventing them from dropping [17]. Recognizing them and taking the proper tutorial are significant in the SPOC education. From this aspect, LSTM and our method are suitable in the situation above, while LSTM are less accurate.

Table 3 Experiments of SBNEDNN and three deep learning methods on Medic 2020

Methods	Acc of L1	Acc of L2	Acc of L3	Acc of L4	Acc of L5	Acc of L6	Tol acc
LSTM	93.06%	100%	97.33%	91.92%	95.65%	82.47%	92.74%
RNN	94.53%	96.15%	88.47%	88.68%	95.21%	78.59%	91.49%
CNN	97.01%	97.11%	98.84%	98.49%	97.82%	84.87%	95.55%
SBNEDNN	95.83%	100%	100%	100%	95.65%	80.41%	**98.85%**

Table 4 Experiments of SBNEDNN and three deep learning methods on Statistics 2020

Methods	Acc of L1	Acc of L2	Acc of L3	Acc of L4	Acc of L5	Acc of L6	Tol acc
LSTM	84.50%	83.66%	96.97%	89.94%	92.77%	52.33%	87.75%
RNN	79.84%	94.77%	94.55%	98.32%	87.35%	69.19%	86.78%
CNN	87.60%	92.81%	96.36%	96.09%	92.77%	79.07%	83.63%
SBNEDNN	96.90%	94.12%	100%	94.41%	98.80%	80.23%	**92.84%**

3.3.2 Model structure comparison

In this section, the prediction performance of different model structures will be compared and discussed. It is divided into position of BN layers and count of fully connected layer. Through these result and relative discussion, it demonstrates that the structure of our model is comprehensive and properly constructed for accurate prediction.

To find the suitable BN layer position, we implemented three positional variants of model in the experiments. The result is shown as Table 5. It is no doubt that SBNEDNN achieves the excellent prediction according to the experimental result, while the network structure without BN layers gets the bottom of experiment as expected. Different positions of BN layer influence the transformation of data to Gaussian distribution, avoiding gradient disappearing and distribution deviation during the propagation. The result also convinces that the BN operation earlier means more positive effect in the network.

Moreover, the amount of fully connected layer plays significant role in the deep learning prediction model. We implemented the modal variants with different multiple fully connected layers. Specifically, all the variant models are BN-embedded. In experiments, we focus on the increasing rate of training time and accuracy, which presents how much time cost in the unit accuracy promotion. Experimental result is presented as Table 6.a and Table 6.b. It demonstrates that more fully connected layers enhance the network accuracy, while it also causes the longer training time and lower adoptability. The result also shows that considerable prediction does not require such deep network which cost much more time. Moreover, the over deep network structure leads to overfitting and data deviation, causing negative effect in the educational predication task. It requires the prediction model with generality and adoptability due to handling the various student behavior features in practical situation.

Table 5 The accuracies with different BN layouts

Structure	Layers	Medic 2020	Statistics 2020
Structure 1	Dense ->Dense ->Dense	67.68%	51.86%
Structure 2	Dense ->Dense ->BN ->Dense	94.83%	62.75%
Structure 3	Dense ->BN ->Dense ->Dense	97.47%	64.21%
SBNEDNN	**Dense ->BN ->Dense ->BN ->Dense**	**98.85%**	**92.84%**

Table 6 Accuracy and training time of three different structures

Layer depth	Medic 2020	Statistics 2020
3	98.85%	92.84%
4	98.92%	93.87%
5	98.99%	95.85%
6	96.74%	94.60%
7	94.52%	91.18%

4 Conclusion

This paper constructs SBNEDNN method for the prediction of SPOC learner's grade. First, an indicator is defined to measure the distribution of dataset as well as guide the sampling process. Second, a BN modified neural network is built to train the data after sampling. Then, a SBNEDNN method is developed to improve the prediction accuracy with imbalance data. Experimental results with the comparison of other three widely used deep learning methods show the effectiveness and supremacy of our method.

Acknowledgments We would like to thank Jiangsu Guidgine Educational Evaluation Inc. and Alliance of Graduate School, Medical University in China for providing SPOC data.

Funding This work was in part supported by the Undergraduate Student Out-of-class Academic Foundation of Nanjing University of Science and Technology, in part supported by the National Natural Science Foundation of China under Grant no. 61801222, and in part supported by the Fundamental Research Funds for the Central Universities under Grant no. 30919011230.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Ahmed M, Tahid SI, Mitu N, Kundu P, Yeasmin S (2020) A Comprehensive Analysis on Undergraduate Student Academic Performance using Feature Selection Techniques on Classification Algorithms. In: *Proc. International Conference on Computing, Communication And Networking Technologies (ICCCNT),* Kharagpur, India
2. Brodic D, Amelio A, Jankovic R (2018) Comparison of different classification techniques in predicting a university course final grade. In: *Proc. International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO),* Opatija, Croatia, pp. 1382–1387
3. Cheng X, Zhang YF, Zhou L, Zheng YH (2020) Visual tracking via auto-encoder pair correlation filter. IEEE Trans Ind Electron 67(4):3288–3297
4. Dalipi F, Imran A, Kastrati Z (2018) MOOC dropout prediction using machine learning techniques: Review and research challenges. In: *Proc. IEEE Global Engineering Education Conference (EDUCON),* Tenerife, Spain, pp. 1007–1014
5. Fu P, Xu Q, Zhang J, Geng L (2019) A noise-resistant superpixel segmentation algorithm for hyperspectral images. Comput Mater Contin 59(2):509–515
6. Hu Q, Rangwala H (2019) Reliable Deep Grade Prediction with Uncertainty Estimation. In: *Proc. International Conference on Learning Analytics & Knowledge (LAK19),* Tempe, AZ, USA, pp. 76–85
7. Ioffe S, Szegedy C (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. [http://arxiv.org/abs/1502.03167 arXiv:1502.03167]. Accessed 26 Oct 2020

8. Juanatas I, Juanatas R (2019) Predictive Data Analytics using Logistic Regression for Licensure Examination Performance. In: *Proc. International Conference on Computational Intelligence and Knowledge Economy (ICCIKE),* Amity University Dubai, UAE, pp. 251–255
9. Liu Q, Liu Y, Liu Y (2018) An integrated framework with feature selection for dropout prediction in massive open online courses. IEEE Access 6:71474–71484
10. Luo H, Pan X, Wang Q, Ye S, Qian Y (2019) Logistic Regression and Random Forest for Effective Imbalanced Classification. In: *Proc. COMPSAC 2021: Intelligent and Resilient Computing for a Collaborative World (COMPSAC),* Milwaukee, WI, USA, pp. 916–917
11. Ran J, Zhang GY, Zheng T, Wang W (2018) Logistic Regression Analysis on Learning Behavior and Learning Effect Based on SPOC Data. In: *Proc. International Conference on Computer Science and Education (ICCSE),* Colombo, Sri Lanka, pp. 330–334
12. Wang X, Yu X, Guo L, Liu F, Xu L (2020) Student Performance Prediction with Short-Term Sequential Campus Behaviors. *Information* 11(4):201
13. Yang T, Brinton C, Joe-Wong C, Chiang M (2017) Behavior-based grade prediction for MOOCs via time series neural networks. IEEE J Sel Top Signal Process 11(5):716–728
14. Yang Y, Fu P, Yang X, Hong H, Zhou D (2020) MOOC learner's final grade prediction based on an improved random forests method. Comput Mater Contin 65(3):2413–2423
15. Zhang G, Sun H, Zheng Y, Xia G, Feng L, Sun Q (2019) Optimal discriminative projection for sparse representation-based classification via Bilevel optimization. IEEE Trans Circuits Syst Video Technol 30(4): 1065–1077
16. Zhang G, Yang J, Zheng Y, Luo Z, Zhang J (2021) Optimal discriminative feature and dictionary learning for image set classification. Inf Sci 574(8):498–513
17. Zhu M, Xia J, Jin X, Yan M, Cai G, Ning G (2018) Class weights random Forest algorithm for processing class imbalanced medical data. IEEE Access 6:4641–4652
18. Zong J, Cui C, Ma Y, Yao L, Chen M, Yin Y (2020) Behavior-driven Student Performance Prediction with Tri-branch Convolutional Neural Network. In: *Proc. The Conference on Information and Knowledge Management '20,* Virtual Event, Ireland, pp. 2353–2356

Affiliations

Zhuonan Liang[1] · Ziheng Liu[1] · Huaze Shi[1] · Yunlong Chen[1] · Yanbing Cai[2] · Hong Hong[3] · Yating Liang[2] · Yafan Feng[1] · Yuqing Yang[4,5] · Jing Zhang[6] · Peng Fu[2]

[1] School of Science, Nanjing University of Science and Technology, Nanjing 210094 Jiangsu, China

[2] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094 Jiangsu, China

[3] Department of Electrical and Computer Engineering, University of California, Davis, Davis, CA 95616, USA

[4] Office of International Cooperation and Exchanges, Nanjing University of Finance & Economics, Nanjing 210046, China

[5] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

[6] Jiangsu Guidgine Educational Evaluation Inc., Nanjing 210046, China