College Students' portrait technology based on hybrid neural network

Zhiming Ding and Xuyang Li

December 25, 2019

College Students' portrait technology based on hybrid neural network*

Zhiming Ding[1] and Xuyang Li[2]

[1] Beijing University of Technology 100022, CN zmding@bjut.edu.cn
[2] Beijing University of Technology 100022, CN lixuyang@emails.bjut.edu.cn

Abstract. Students have produced a large number of data in the teaching life of colleges and universities. At present, the development trend of university data is to gradually form a high-dimensional data storage system composed of student status information, educational administration information, behavior information, etc. It is of great significance to make use of the existing data of students in Colleges and universities to carry out deep-seated and personalized data mining for college education decision-making, implementation of education and teaching programs, and evaluation of education and teaching. Student portrait is the extension of user portrait in the application of education data mining. According to the data of students' behavior in school, a labeled student model is abstracted. To address above problems, a hybrid neural network model is designed and implemented to mine the data of college students and build their portraits, so as to help students' academic development and improve the quality of college teaching. In this paper, experiments are carried out on real datasets (the basic data of a college's students in Beijing and the behavior data in the second half of 2018-2019 academic year) . The results show that the hybrid neural network model is effective.

Keywords: Big data · Data mining · Higher education · Convolutional neural network · Feedforward neural network.

1 Introduction

In the past few years, revolutionary changes have occurred in the field of education and information. Online learning systems, smartphone applications and social networks have provided a large number of applications and data for educational data mining (EDM) research. Massive open online courses (MOOCs) became a new type of teaching model in the recent 2 years. It's obviously that EDM is in the era of "big data", which tells that EDM research will be developed rapidly. [1]

As far as the current situation is concerned, it is very meaningful to carry out a deep-level and personalized data mining for the existing student data of

* Supported by the National Key R&D Program of China under grant number 2017YFC0803300.

colleges and universities. Student portraits are an extension of user portraits in educational data mining applications, which makes an abstract label student model according to the data of students' behavior in school.

The establishment of a student portrait system in colleges and universities has the following significance: the adjustment of educational decision-making has avoided problems such as partial models and inaccurate data due to the lack of analysis tools; the correct implementation of education and teaching programs has avoided large differences due to students In the management of students, attention is paid to the classification and partitioning and dynamic adjustment, which affects the teaching effect; to achieve efficient management of students, real-time understanding of students' thinking dynamics, and accurate management of students.

At present, many universities collect and display the relevant data generated by students' work, instead of research on data mining. In terms of student portraits, their research is more about generating data for students The statistics on the number and frequency of information. The portrait tags are more derived from relevant higher education researchers based on experience and are not comprehensively representative.

Therefore, this paper aims at the basis of hybrid neural networks, based on some basic information (birthplace, ethnicity, gender, etc.) currently known to undergraduates in a certain college and university, the educational administration information data generated by students during the semester, and the behavior information data generated by campus card and gateway account on the basis of hybrid neural network for students to help students academic development and to improve the teaching quality of colleges and universities.

2 Related work

During the preliminary investigation, we found that the current research on EDM is not enough in universities, and most of them stay on the statistics, collation and summary of student data, or use relatively simple machine learning methods, such as decision tree classifiers and naive Bayes classifiers [2, 3] , it is impossible to predict and analyze the potential problems or tendencies of students. In horizontal comparisons in other industries, we found that a large number of deep learning algorithms have been applied to user data mining in industries such as e-commerce, socialization, and finance [4–6] . And text classification, we learned from this and optimized it, and proposed a portrait technology for college students based on hybrid neural networks.

2.1 Portrait Technology

The concept of "user portrait" was first proposed by Alan Cooper, the father of interaction design [7]. It refers to an abstract labelled user model. This model is based on basic user information, social information, preference information and behavior. The information is summarized. In the process of forming user portraits, the most core step is to apply appropriate "tags" to the users. These labels to the portraits are generated by analyzing the collected user data, and can be used for these The data information is highly summarized.

User portraits are also called user roles. As an effective tool for sketching target users, connecting user demands and design directions, user portraits have been widely used in various fields. We often use the most obvious and close word to life in the actual operation process to connects user attributes, behaviors and expectations. As a virtual representative of actual users, the user roles formed by user portraits are not constructed outside the product and the market. The user roles formed need to be representative and capable of Represent the main audience and target group of the product.

Li [8] and others combined the 38 million e-commerce activity data generated by Groupon with the combination of domain knowledge from e-commerce with data mining and graph theory methods to improve the user viscosity of the product; Lu [9] proposed an algorithm for topic interest mining of Sina Weibo users based on tags and two-way interactions. The algorithm for mining user topic interests through free tags and social interaction on Sina Weibo. Wikipedia expanded and expressed as vectors, and uses the tf-idf method to obtain the vector of each microblog. Then analyzes the sort of user interest tags according to the forward and the backward arrangement to get the final user interest tags; Punit [10] proposed a hybrid, set-based clustering algorithm that can use fast data space reduction and smart sampling strategies; Liu [11] proposed an improved clustering-based collaborative filtering recommendation algorithm introduces time decay function for preprocessing the user's rating and uses project attribute vectors to characterize projects, user interest vectors to users and use clustering algorithms to cluster the users and the projects respectively. The algorithm can portrait for users in multi dimension and reflect the user's interest changing. Jia [12] employed the telecom data and proposes a user clustering and influence power ranking scheme. The scheme is implemented through three stages, i.e. the user portrait analysis stage, the user clustering analysis stage and the ranking stage of user influence power; Cuiling [13] proposed a new multi-dimensional foreign language learning community, which dynamically evaluates and classifies students' learning status in real time to form student portraits, and automatically guides students to perform group learning at different levels of ability in the learning community. Choosing a reasonable and effective clustering algorithm successfully implemented the system's data mining application in the e-learning environment.

In the study of user portraits, we learned that generating new labels by clustering algorithms is a very common method in corporate user portraits, and the current user portrait technology is not suitable for college data. There are differences on purpose of college user portrait and enterprise user portrait. Student portrait is the production of college big data, which is applied to help administrator of college to know students, master student status by real-time and exactly locate abnormal crowed. Otherwise, enterprise portrait is more applied to deal with propensity to consume and propensity of topic.

From the current research on student portraits in universities, we can see that there are two major problems with the existing student portrait systems. On the one hand, many studies are only based on the system architecture of student portraits. Called for discussion without mentioning specific student portrait establishment methods; on the other hand, the existing student portrait system analysis has fewer dimensions, the breadth of the data source for analysis is not sufficient, and other dimensions of the student's information, such as behavior information, etc. Synthesize student portraits. In addition, the method of establishing student portrait tags is more based on the experience of higher education researchers. It does not have comprehensive representation and the data mining model is too simple to reflect the recent thinking of students in real time. Dynamic, and model accuracy is not high.

2.2 Neural network behavior prediction

Donkers [14] extended the regression neural network by considering the unique features of the recommendation system domain. One of these features is a clear concept of user recommendations specifically generated for it, showing how to use a sequence of consumer items in a new type of gating loop unit. To represent individual users to effectively generate personalized recommendations for the next project; Lefebvre-Brossard [15] proposed a new method that relies on recurrent neural networks and word embeddings to match the problem of learners looking for guidance, and mentors willing to provide such guidance; Wang [16] proposed an end-to-end encrypted traffic classification method with one-dimensional convolution neural networks. This method integrates feature extraction, feature selection and classifier into a unified end-to-end framework, intending to automatically learning nonlinear relationship between raw input and expected output; Lin [17] designed a convolutional neural network with cross autoencoders to generate user-scope content attributes from low-level content attributes. Finally, they propose a deep neural network model to incorporate the two types of userscope attributes to detect users' psychological stress. Tommy [18] used one-dimensional convolution neural network and short-term memory neural network to mine user information on Facebook, so as to predict user personality. Cai [19] proposed a CNN-LSTM attention model to predict user intents, and an unsupervised clustering method is applied to mine user intent taxonomy.

From the above, we can see that when using deep learning to predict user behavior, convolutional neural networks and recurrent neural networks are often used. The advantage of convolutional neural networks is that they can store data in high dimensions and can introduce the time axis is used as the data dimension to solve the complexity of the data structure of college students. The advantage of the recurrent neural network is that the data itself has a requirement for the input sequence. It has a strong time correlation and can better connect students with Time is used as the dimension data during the school, such as teaching week, but in actual operation, we found that the current student data of colleges and universities cannot evaluate students in a short unit time. One week's performance is evaluated, so it is impossible to implement classification prediction of student portrait tags with RNN, so we chose to use CNN as a component of the model for classification and prediction of portrait tags when building a hybrid neural network model.

3 Label design of student portrait based on high dimensional clustering

3.1 clique algorithm

Students in colleges and universities will generate all kinds of data, such as student basic data, teaching data, card consumption data, access control data, library borrowing data, gateway traffic, web browsing data, etc. These data can be divided into structures Data and unstructured data. Structured data is structured data. Generally, the amount of data is small, but it has good data characteristics. Unstructured data includes semi-structured data and unstructured data. The amount of data is huge and the characteristics of the data are not obvious. Using high-dimensional clustering technology, we can perform data mining, transform complex multi-source heterogeneous data into simple semantic labels, and lay the foundation for the establishment of student portraits.

The grid clustering algorithm has good scalability for data set size, can handle large-scale data sets, and the clustering results are not affected by the order of data input. It is suitable for the diverse, high-dimensional, and individual student data sources of college students in this paper. The characteristics of the grid clustering algorithm are more intuitive and easy to understand, and it can be used to classify student data in the management of college students in a practical way, and positively promote higher education research.

This paper plans to use clique-based clustering algorithm, that is, automatic subspace clustering algorithm, to cluster students' various data, and to explore the potential connections between students in static data and dynamic data, so as to establish student portrait labels. The model is more suitable to explore the degree of influence between various attributes of students, so as to determine the impact of specific attributes or behaviors on students.

Clique has the advantages of grid-like algorithms and is not sensitive to the order of data input. It does not need to assume any standardized data distribution. It scales linearly with the size of the input data. It has good scalability when the data dimension increases. For large databases The clustering of high-dimensional data in is very effective. The advantages of the above clique are very suitable for dealing with student data, the input of student historical data does not require order, and the data structure is highly designable, which is easy to expand and adjust.

The clique algorithm pseudo-code is shown in Algorithm 1.

The clique algorithm uses a fixed mesh division method to divide the data space equally according to different parameters entered by the user, and this will cause some data sample points in the sparse area at the class boundary that should belong to this clustering cluster to be considered non-compliant

Algorithm 1. The clique algorithm	
Input:	D_{k-1} (Set of all $k-1$ dimensional dense units);
Ouput:	all k-dimensional candidate denses;
Algorithm:	
	insert into S_k ;
	select $u_l[l_1, h_1], u_l[l_2, h_2], ..., u_l[l_{k-1}, h_{k-1}], u_l[l_{k-2}, h_{k-2}]$;
	from $D_{k-1}u_1, D_{k-2}u_2$;
	where $u_1a_1 = u_2a_1, u_1l_1 = u_2l_1, u_1h_1 = u_2h_1, u_1a_1 = u_2a_2, u_1l_2 = u_2l_2,$
	$u_1h_2 = u_2h_2, u_1a_{k-2} = u_2a_{k-2}, u_1l_2 = u_2l_2, u_1h_2 = u_2h_2, ...,$
	$u_1a_{k-2} = u_2a_{k-2}, u_1l_{k-2} = u_2l_{k-2}, u_1h_{k-2} = u_2h_{k-2}, u_1a_{k-1} < u_2a_{k-1}$;

Dense units may be partitioned into different adjacent sections, which seriously damages the integrity of the original dense area.

According to the strategy of pruning the minimum description length, the entire data space is divided into multiple different subspaces. In each subspace, the density units are forcibly divided into n groups and the corresponding data is covered. Under this rule, the largest one can be found. Subspace, the remaining subspaces will be ignored. Under this strategy, some of the subspaces have been pruned, and the dense units you are looking for may be located in it, which leads to the incompleteness of the dense units. Although the method improves the operation efficiency of the algorithm, the lost subspaces affect the accuracy of the clustering results.

3.2 Auto-CLIQUE algorithm

In high-dimensional data space, the number of grid cells will increase exponentially, and fixed grid division based on input parameters will likely cause the same cluster to be divided into multiple regions. The proposed discrete coefficients adopt adaptive meshing based on discrete coefficients and set iterative grid density thresholds to dynamically and flexibly divide the mesh to improve the processing effect and performance of the algorithm without the need for user input. Reliance on the knowledge of experiments to improve the quality of clustering.

Adaptive Meshing Based on Discrete Coefficients In order to eliminate the influence of the level of data value and different measurement units on the measurement of the degree of dispersion, a dispersion coefficient is set, which is the ratio of the standard deviation of a set of data to its mean value. It is used to measure the relative degree of dispersion of the data, and does not require user input Parameters, reducing the algorithm's dependence on prior knowledge and improving the quality of clustering. The formula for the discrete coefficient is:

$$D_j = \frac{S_j}{\overline{X}_j} \tag{1}$$

where D_j is the dispersion of the j-dimensional data of the data set, S_j is the standard deviation of the j-th dimension, $\overline{X}_j$ is the mean of the j-dimensional data. The standardized dispersion formula is:

$$D_s = \frac{1}{1 + \frac{1}{d} * \sum_{d}^{j=1} D_j} \tag{2}$$

where D is the dimension of the data object, D_s the smaller the value of, the greater the dispersion of the data set, D_s the larger the value of, the smaller the discreteness of the data set. The definition of the segmentation parameter m is:

$$m = D_s * \sqrt[d]{N} \tag{3}$$

where N is the number of samples in the data set.

Grid density threshold setting based on iteration In addition, we also propose to use a recursive algorithm to calculate the data density threshold according to the data set itself, which can effectively separate the dense grid and the sparse grid, and once again reduce the dependence on prior knowledge. The density threshold formula is:

$$\rho[i+1] = \rho[i] + \frac{grid[i]}{data[i]} \tag{4}$$

where $\rho[i]$ is the ith density threshold, $grid[i]$ is greater than or equal to the density threshold $\rho[i]$ the number of grids, $data[i]$ for $grid[i]$ grids exceed density threshold $\rho[i]$ the total number of extra data points.

4 Classification model of student portraits based on hybrid neural network

The recent labeling of student portraits is more time-efficient than the feature labeling a long time ago, but at the same time the meaning of labels generated by static data cannot be denied, so this paper plans to propose a method based on recurrent neural networks and feed forward neural networks. Hybrid neural network models to solve this situation.

4.1 Hybrid neural network framework

In the process of establishing the student portrait label, we found that taking the student's "learning effort" label as an example, the label is not only affected by the performance results of the teaching affairs data in the static data, but also by the student's routine of the semester. The frequency and other behavioral effects of entering and leaving a school, that is, the labeling results are affected by both the long-term static properties of the students and the short-term behaviors. The two results must have different choices for the labeling results.

In this paper, the following improved one-dimensional convolutional neural network and BP pre-feedback neural network are combined with SotfMax and trained with real data. The network structure is shown in Fig.1.

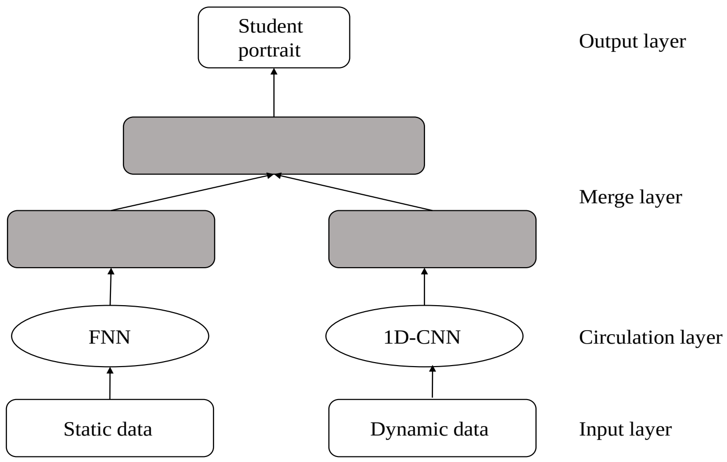


Fig. 1. Basic framework of hybrid neural network

4.2 Data prediction based on feed forward neural network

Compared with the student's dynamic behavior data, the static data composed of the student's basic information and teaching information is much easier to handle, but it represents the student's long-term behavioral trend, so we can use this kind of data to predict the student's long-term development trend. This paper uses a front-feedback neural network (FNN) to perform data mining on this part of the data. The BP neural network is used to determine the corresponding labels of the student portraits. An example is shown in Fig.2.

We use a vector data structure to store training data and test data in the BP neural network. In addition to the basic information data of the students, it also includes relevant data about the last semester about campus life in the campus: average dining hall consumption , Standard deviation, number of consumptions; average consumption, standard deviation, and number of times of consumption in supermarkets; average consumption, standard deviation, and number of times of consumption in online schools; average number of consumption, averages of bathroom recharges in schools, and educational data since students enrolled: all courses Average scores, weighted scores, and standard deviations of scores, as well as the average scores of public basic compulsory courses, practical compulsory courses, general education and school elective courses, subject basic compulsory courses, subject basic elective courses, professional optional courses, and pro-

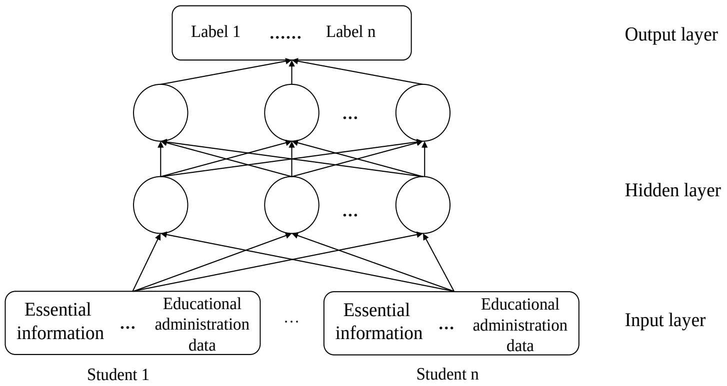


Fig. 2. Basic framework of hybrid neural network

fessional limited elective courses , Weighted score, standard deviation of scores. The data of each student is shown in Table 1 .

Table 1. FNN student data sheet

Type of data	Data breakdown	Data field
Living data	Canteen consumption	Mean, standard deviation, Consumption
	Supermarket Consumption	Mean, standard deviation, Consumption
	Recharge the school bathroom	Consumption times, average
Academic data	All courses Public Foundation Compulsory Course Practical Compulsory Course General Education and School Electives Required Subject Courses Subject Elective Courses Professional optional Courses Restricted Courses	Average score, Weighted score, Fractional standard deviation
Basic information data		

To ensure accuracy, we set up four-layer neurons for the BP neural network in this paper. The activation function uses the relu activation function mentioned above.The network structure is shown in Table 2.

Dynamic data prediction based on one-dimensional convolutional neural network (1D-CNN) Compared to the static data of students, the dynamic data of students has a lot of time-oriented behavioral data, which is more complicated. In it, we find that the dynamic data of students is different from the

Table 2. FNN data structure table

Layer(type)	Output Shape	Param #
Input	(None, 26)	0
Dense1	(None, 108)	2916
Dense2	(None, 52)	5668
Dense3	(None, 12)	636
Dense4	(None, 6)	78

static data in two data formats, which are integrated into The unified data format will cause a large amount of compression and loss of data, or there will be a large amount of data redundancy. For student behavior data, we clean and organize the data and store it in a matrix. The data is as follows as shown in Table 3.

Table 3. 1D-CNN student data sheet

Student serial number	Card data item	Time (teaching week)			
Student n	Canteen Consumption Canteen consumption standard deviation Total supermarket consumption Supermarket consumption standard deviation Total online consumption Total bathroom consumption	First 1 week	First 2 week	...	First 20 week

Dynamic data exists in the form of a two-dimensional matrix. In deep learning, convolutional neural networks have always studied in depth two-dimensional data such as video, images, and audio. One-dimensional convolutional neural networks are suitable for sequence data or languages. Data, so we borrowed its way of processing two-dimensional data and used one-dimensional convolutional neural networks to process student dynamic data.

The structure of a common one-dimensional convolutional neural network model is: an input layer, a convolutional layer, a pooling layer, a three-layer fully connected layer, and an output layer. The convolutional layer can be expressed as:

$$C = f(xk + b) \tag{5}$$

where x represents the input, k represents the convolution kernel, and b represents the offset value. f is the activation function. Common activation functions include relu, tanh, and sigmoid. The relu function is used in this paper (see equation (6)) .

$$f(x) = max(0, x) \tag{6}$$

The convolution layer C performs sliding on the serialized data set and convolves with the original data to obtain the feature layer.

Pooling layer S, the pooling layer refers to the down sampling layer, which combines the output of a cluster of neurons in the previous layer with a single neuron in the lower layer. The pooling operation is performed after non-linear activation, where the pooling layer helps reducing the number of parameters and avoiding over fitting, it can also be used as a smoothing method to eliminate unwanted noise. The S layer can be expressed as:

$$S = \beta \cdot (C) + b \tag{7}$$

where β and b are scalar parameters, and *down* is a function selected by down sampling. There are an average pooling layer and a maximum pooling layer. This paper uses both methods to improve the functionality of the perception area.

The output layer uses a $Softmax$ function classifier, assuming the output of a one-dimensional convolutional neural network is $y_1, y_2, ..., y_n$, and the output after $Softmax$ layer is:

$$Softmax\left(y\right)_i = \frac{e^{y_i}}{\sum_n^{i=1} e^{y_j}} \tag{8}$$

The one-dimensional convolutional neural network model designed according to the data mining needs of college students in this paper is shown in Fig.3.

The data structure of each layer in 1D-CNN is shown in Table 4.

Table 4. 1D-CNN data structure table

Layer(type)	Output Shape	Param #
Reshape	(None, 18, 8)	0
Conv1D	(None 16, 40)	1000
MaxPooling1D	(None, 8, 40)	0
LSTM1	(None, 8, 80)	38720
LSTM2	(None, 8, 20)	8080
AveragePooling1d	(None, 20)	0
Dense	(None, N)	42

Between the largest pooling layer and the average pooling layer, we have introduced two layers of long-term and short-term memory networks (LSTM). Using the recurrent neural network's memory and parameter sharing features, it improves the connection between the student portrait label and the upper and lower teaching week. The relevance of making predictions.

The LSTM network is a time-recurrent neural network, which is specially designed to solve the long-term dependency problem of general RNN (recurrent neural network).

In the text $x_1, x_1, ..., x_n$ represents the output of the previous MaxPooling layer, with a structure of 40 channels in 1 row and 8 columns. x_n corresponds

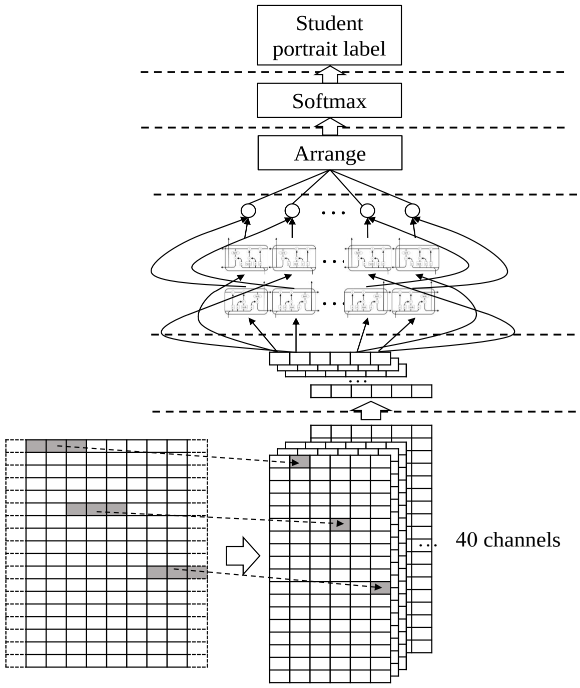


Fig. 3. Improved 1D-CNN model structure

to the n-th column (n maximum 8) array, this array includes 40 eigenvalues, $h_1, h_1, ..., h_n$ express $x_1, x_1, ..., x_n$ after each output of the cell, when the LSTM structure is 80 channels in 1 row and 8 columns, h_n express x_n (40 eigenvalues) after the cell output (80 eigenvalues). A schematic reference is shown in Fig.4.

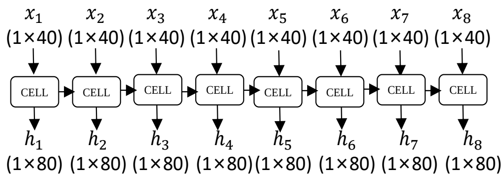


Fig. 4. LSTM layer data structure in 1D-CNN

The training objective function uses a cross-entropy formula function with the following formula:

$$L = -[ylog\hat{y} + (1 - y)\log(1 - \hat{y})] \tag{9}$$

In equation (9), when $y = 1$,$L = -\log\hat{y}$ the closer the predicted output is to the true sample label 1, the smaller the loss function L is; The closer the predicted output is to 0, the larger L is. Therefore, the change trend of the function completely meets the actual needs; when $y = 0, L = -\log(1 - \hat{y})$ the closer the predicted output is to the true sample label 0, the smaller the loss function l; The closer the predicted function is to 1, the larger l, the change trend of the function also fully meets the actual needs.

5 Experiment

5.1 Data set and experimental environment

Window version: Windows 10 64-bit operating system Hardware environment: i7-6700HQ CPU 2.60GHz; RAM 8.00GB Anaconda3 running environment: Python 3.6.0; Python Package numpy 1.15.2, pandas 0.25.1, tensorflow 1.14.0, keras 2.24, scikit-learn 0.21.3, matplotlib 3.1.1, seaborn 0.9.0.

5.2 Model evaluation criteria

Confusion matrix As shown in Table 5, for the prediction of obtaining scholarship, the examples can be divided into true examples (True Positive (TP)), false positives (FP), true Negative (TN), and False Negative (FN) according to the combination of their real categories and algorithm recognition. We combine the above to form a "confuse matrix of predicting scholarship".

The true example is the correct choice; the false positive example is the wrong choice, which indicates a misjudgment, and the accuracy of the result is related;

Table 5. Confusion matrix of predicting scholarship

Reality	Forecast result	
	obtain scholarship	not acquired scholarship
obtain scholarship	TP (real example)	FN (false counterexample)
not acquired scholarship	FP (false positive)	TN (true and negative)

the false negative example indicates the missing data, which is related to the recall rate; the true and negative examples are not needed in this paper and will not be discussed.

Equation (10) represents the accuracy rate. The ratio of the number of TP in the recognition results found in a single run of the algorithm is as follows:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{all\ detections} \tag{10}$$

Accuracy, recall and F1-Score indicators Equation (11) represents the recall rate. The ratio of the number of TP found in a single run of the algorithm to the number of all positive samples in the sample. The expression is:

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{all\ ground\ truths} \tag{11}$$

Equation (12) represents the F1-Score indicator, which combines the results of $Precision$ and $Recall$. The value of $F1 - Score$ ranges from 0 to 1, where 1 represents the best model output, and 0 represents the model output result, worst. The expression is:

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{12}$$

5.3 Experimental results and analysis

With regard to the hybrid neural network model mentioned above, we selected 1,055 undergraduates in a certain college and university to train the model, which can be used to determine whether there is an academic crisis for students, whether they are potentially poor students, whether there are psychological warnings, It is expected that scholarships will be obtained and recent graduates will become outstanding graduates.

BP neural network model test Taking the expected scholarship for students as an example, the results of the cross-entropy loss function during the training of the BP neural network model using the corresponding data are shown in Fig.5.

The related test parameters of the BP neural network model are shown in Table 6.

Fig. 5. Cross entropy loss function in the training process of BP neural network model

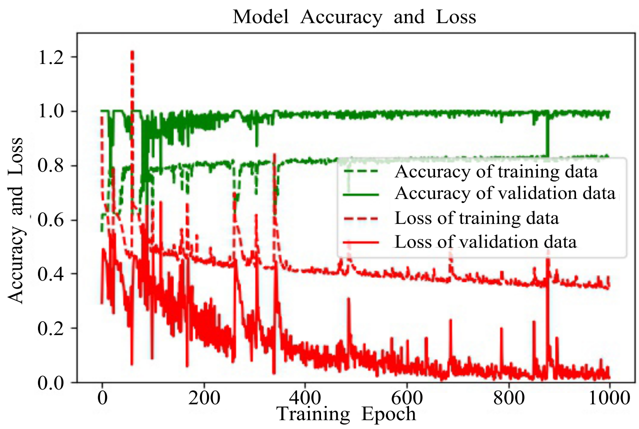


Table 6. Comparison of four neural networks validation parameters

Type	Data type		Precision		Recall		F1-score		Time	Accuracy
	static	dynamic	0	1	0	1	0	1	(us/step)	
Common 1D-CNN		√	0.70	0.41	0.85	0.66	0.75	0.24	42	0.66
Improved 1D-CNN		√	0.73	0.40	0.72	0.41	0.72	0.41	126	0.62
BP neural network	√		0.86	0.83	0.94	0.67	0.90	0.74	27	0.85
Hybrid neural network	√	√	0.88	0.78	0.91	0.74	0.89	0.76	160	0.85

ps: 0 in the table means no scholarship (722 supports) , 1 means scholarship (333 supports)

From the chart above, we can see that the *accuracy* of the BP neural network model in this paper is 0.85, the $F1-score$ of the un-scholarship is 0.90, and the F1-score of the scholarship is 0.74. The results of scholarships are obtained for prediction. The main reason for the better performance of this model is that the training data contains the student's various achievements, which can affect the possibility of students receiving scholarships to a large extent.

Improved 1D-CNN neural network model test In the improved 1d-cnn neural network model test, we used a common 1d-cnn neural network model with a similar structure to compare with it. In addition, we implement it according to the one-dimensional neural network model proposed by Waibel A [20] , and compare it with other neural network models in this paper. The network structure is shown in Table 7.

Table 7. Common 1D-cnn data structure table

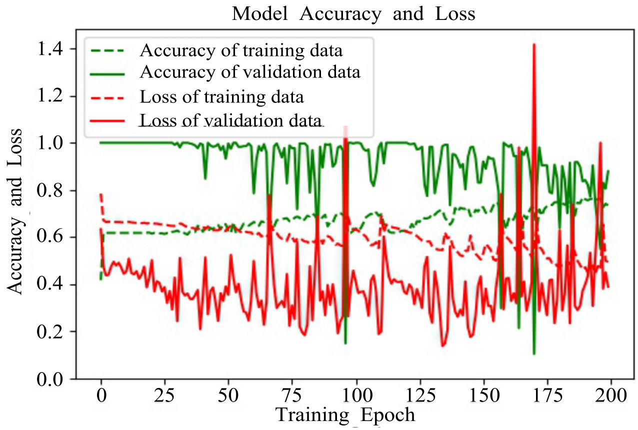
Layer(type)	Output Shape	Param #
Reshape	(None,18,8)	0
Conv1D	(None,16,40)	1000
MaxPooling1D	(None,8,40)	0
AveragePooling1D	(None,20)	0
Dense	(None,N)	42

This section of the experiment uses the school life data of the students mentioned in the 2018-2019 school year as input, and each student's data is stored in a matrix in the form of Table 3. The parameter settings are as follows: the kernel size is set to 3, Use the $EarlyStopping$ function as a callback function, set vac_loss to monitor, and set the value of the patience parameter to 200, which means that during the training process, when the loss value of the test set is constant for 200 trainings, stop training in advance. Training batch size (batch size) is set to 256, the learning rate is set to 0.01, and the maximum number of iterations (epoch) in the training phase is set to 2000.

The two 1D-CNN model pairs are shown in Table 6. From Table 6, we can see that the improved 1D-CNN model has improved the classification accuracy rate, classification recall rate, $F1-Score$, and *accuracy* rate of whether students will receive scholarships, especially for "getting scholarships" Relevant parameters of classification, because in the data set, the number of students corresponding to the "not awarded scholarship" label is higher than "received scholarship", so the improvement of this type of parameters indicates that the improved 1D-CNN model is more common than the ordinary 1D-CNN model The accuracy has been greatly improved. However, as the improved 1D-CNN model adds two LSTM layers to enhance the connection of student life data during the teaching week, the time complexity has increased. It takes time from 42 us to 126 us.

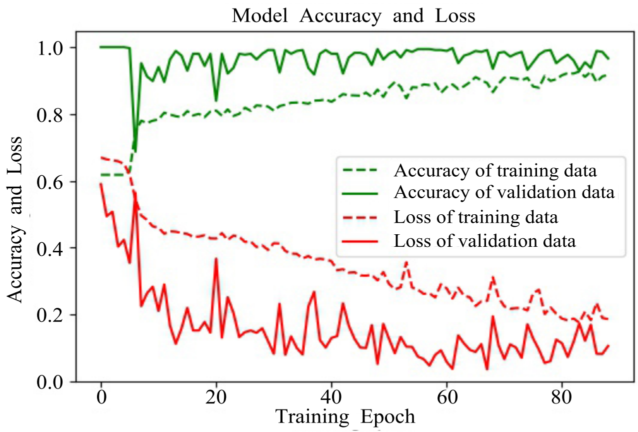
The results of the cross-entropy loss function training using the improved 1d-cnn model in this paper are shown in Fig.6.

Fig. 6. Cross entropy loss function in the training process of improved 1D-CNN model



Hybrid neural network model test The cross entropy loss function during the training of the hybrid neural network model mentioned in this paper is shown in the Fig.7. The overall trend is that the accuracy of the training data and test data gradually approaches 1, and the loss gradually decreases. The training convergence process realizes the connection of static and dynamic data of students, and can mine the potential relationship between static data such as student performance in various subjects and dynamic data generated from campus life, and can better predict the situation of students receiving scholarships. Provide guidance for college student management.

Fig. 7. Cross entropy loss function in the training process of hybrid neural network model



The relevant test values are shown in Table 6. From the table above, we can see that in the hybrid neural network model, students have higher accuracy in the prediction of not receiving the scholarship label, and have certain prediction capabilities in obtaining the scholarship label, and the overall accuracy can be correct. Predicting whether students will receive scholarships has achieved our goal of building a hybrid neural network.

6 Conclusion

This paper builds a college student portrait system based on an improved one-dimensional convolutional neural network and a front-feedback neural network. Among them, an optimization is proposed in the establishment of a one-dimensional convolutional neural network model. The 1d-cnn maximum pooling layer, the LSTM layer and the average pooling layer strengthen the connection of college student data in the unit of teaching week.

In addition, more information is provided on student personal data mining, predicting students' future behavioral tendencies, so as to implement precision poverty alleviation, academic early warning, psychological early warning and other tasks in efficient work.

The future development direction is to implement the application of algorithms based on the specific status of universities, establish a large database of student behavior information, establish a student information IoT acquisition system, and establish a student information distributed system, so as to improve data collection, data cleaning, and real-time prediction functions.

References

1. Zhou Q, Mou C, Yang D. Research progress on educational data mining: A survey. Journal of Software[J] 2015,26(11):3026−3042 (in Chinese).
2. Castillo G , João Gama, Breda A M . Adaptive Bayes for a Student Modeling Prediction Task Based on Learning Styles[C]// User Modeling 2003, 9th International Conference, UM 2003, Johnstown, PA, USA, June 22-26, 2003, Proceedings. Springer Berlin Heidelberg, 2003.
3. Pandey M , Sharma V K . A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction[J]. International Journal of Computer Applications, 2013, 61(13):1-5.
4. Yuan H , Xu W , Wang M . Can online user behavior improve the performance of sales prediction in E-commerce?[C]// IEEE International Conference on Systems. IEEE, 2014.
5. Alahi A , Goel K , Ramanathan V , et al. Social LSTM: Human Trajectory Prediction in Crowded Spaces[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
6. Wong B K , Selvi Y . Neural network applications in finance: A review and analysis of literature (1990–1996)[J]. Information and Management, 1998, 34(3):129-139.
7. Brickey J , Walczak S , Burgess T . Comparing Semi-Automated Clustering Methods for Persona Development[J]. IEEE Transactions on Software Engineering, 2012, 38(3):537-546.

8. Li K , Deolalikar V , Pradhan N . Mining lifestyle personas at scale in e-commerce.[C]// IEEE International Conference on Big Data. IEEE, 2015.
9. Deng L , Jia Y , Zhou B , et al. User interest mining via tags and bidirectional interactions on Sina Weibo[J]. World Wide Web-internet & Web Information Systems, 2018, 21(2):515-536.
10. Punit R , Dheeraj K , Bezdek J C , et al. A Rapid Hybrid Clustering Algorithm for Large Volumes of High Dimensional Data[J]. IEEE Transactions on Knowledge and Data Engineering, 2018:1-1.
11. Xiaojun, Liu. An improved clustering-based collaborative filtering recommendation algorithm[J]. Cluster Computing, 2017, 20(2):1281-1288.
12. Jia Y , Chao K , Cheng X , et al. Big data based user clustering and influence power ranking[C]// 2016 16th International Symposium on Communications and Information Technologies (ISCIT). IEEE, 2016.
13. Lv C . Application Study on Data Mining Technology of English Learning Virtual Community[C]// International Conference on Intelligent Transportation. IEEE Computer Society, 2018.
14. Donkers T , Loepp B , Jürgen Ziegler. Sequential User-based Recurrent Neural Network Recommendations[C]// Eleventh Acm Conference on Recommender Systems. ACM, 2017.
15. Lefebvre-Brossard A , Spaeth A , Desmarais M C . Encoding User as More Than the Sum of Their Parts: Recurrent Neural Networks and Word Embedding for People-to-people Recommendation[C]// Conference on User Modeling. ACM, 2017.
16. Wang W , Zhu M , Wang J , et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks[C]// 2017 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, 2017.
17. Lin H , Jia J , Guo Q , et al. User-level psychological stress detection from social media using deep neural network[J]. 2014.
18. Tommy Tandera, Hendro, Derwin Suhartono, et al. Personality Prediction System from Facebook Users[J]. Procedia Computer Science, 2017, 116:604-611.
19. Cai R , Zhu B , Ji L , et al. An CNN-LSTM Attention Approach to Understanding User Query Intent from Online Health Communities[C]// 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2017.
20. Waibel A , Hanazawa T , Hinton G , et al. Phoneme recognition using time-delay neural networks[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 2002, 37(3):328-339.