

Research Article

Analysis of College Students' Public Opinion Based on Machine Learning and Evolutionary Algorithm

Jinqing Zhang ¹, Pengchao Zhang ², and Bin Xu ^{3,4}

¹School of Automation Science and Electrical Engineering, Beihang University, Beijing, China

²Shaanxi Provincial Key Laboratory of Industrial Automation, Shaanxi University of Technology, Hanzhong, Shaanxi 723000, China

³Major Public Information Research Center of Shaanxi Province, Northwestern Polytechnical University, Xi'an, China

⁴School of Marxism, Northwestern Polytechnical University, Xi'an, China

Correspondence should be addressed to Bin Xu; smileface.binxu@gmail.com

Received 21 April 2019; Accepted 9 September 2019; Published 11 November 2019

Guest Editor: Gonzalo Farias

Copyright © 2019 Jinqing Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The recent information explosion may have many negative impacts on college students, such as distraction from learning and addiction to meaningless and fake news. To avoid these phenomena, it is necessary to verify the students' state of mind and give them appropriate guidance. However, many peculiarities, including subject focused, multiaspect, and low consistency on different samples' interests, bring great challenges while leveraging the mainstream opinion mining method. To solve this problem, this paper proposes a new way by using a questionnaire which covers most aspects of a student's life to collect comprehensive information and feed the information into a neural network. With reliable prediction on students' state of mind and awareness of feature importance, colleges can give students guidance associated with their own experience and make macroscopic policies more effective. A pipeline is proposed to relieve overfitting during the collected information training. First, the singular value decomposition is used in pretreatment of data set which includes outlier detection and dimension reduction. Then, the genetic algorithm is introduced in the training process to find the proper initial parameters of network, and in this way, it can prevent the network from falling into the local minimum. A method of calculating the importance of students' features is also proposed. The experiment result shows that the new pipeline works well, and the predictor has high accuracy on predicting fresh samples. The design procedure and the prediction design will provide suggestions to deal with students' state of mind and the college's public opinion.

1. Introduction

Youth is the most important period for college students to establish a mature outlook on life and values. In college, students' perception on life and various things includes public opinion, which can also influence the ideology of students in turn. The advent of Internet has increased the diversification of mass media, which makes it possible for people to obtain information that they are interested in at anytime and anywhere. However, the quality and reliability of information show increasing difference. Some untrue and negative information might pollute public opinion in college and cause harmful influence on students' state of mind. For personality, research studies have shown that students who

are addicted to Internet and wireless mobile devices such as smartphones relate to increase in stress and anxiety while decrease in academic performance and satisfaction with life [1, 2]. These impacts could make students take a pessimistic view and feel their lives meaningless which show strong relationship with depressive disorder and even suicide. For society, the spread of rumors could make students more suspicious and treat social media and government as liars instead [3]. When students enter society after graduation, their distrust on government will leave room to disharmony. A student's state of mind is the cell of public opinion in college, and there have been strong evidences showing that students in positive environments are more likely to make great achievements [4]. To protect students from the

negative impact of information explosion, colleges should focus on giving guidance to students with problems in mind, take responsibility for helping them correct their outlook on life and values, and make them be willing to fight for the development of the whole human race.

However, students are usually not willing to seek guidance on state of mind because many of them do not want to be regarded as “sick.” This requires the colleges to actively implement guidance on students. But if students tend to hide their problem, there will be problems for colleges to know who needs to be guided when facing thousands of students. One of the methods is using machine learning (ML) tools such as the neural network (NN) to predict students’ state of mind. ML tools can automatically learn the function from students’ features to their state of mind and make prediction quickly and accurately as long as there are enough training data. With precise prediction on students’ state of mind, colleges can adjust the guidance according to the students’ own features to enhance its effectiveness [5].

ML has been widely used to predict people’s opinion on things by doing text analysis on data collected from Internet, but it might not be so much useful when predicting students’ state of mind. That is because prediction of student’s state of mind has several peculiarities: (1) focus on subject: this work is focused on the people who make judgements, but not the judgements they have made; (2) multiaspect: to enhance the correctness of the analysis, the predictor should learn plenty of information from different aspects, but students might not publish some of this information forwardly on the Internet; (3) low consistency on aspects: different students would like to pay attention to different matters, thus it is opinionated to make an answer on a certain question as a public criterion. To meet these peculiarities, more abundant data should be collected for a single sample which covers most aspects of opinions related to a student’s daily life, and the data of different samples should have good consistency on their content. If only text-based data from the internet are collected, the data set will be not effective enough. On the contrary, the traditional method of using the questionnaire to get the data can better meet the requirements. The questionnaire used is well designed to cover most of the aspects about college students, and the questions with scale can help quantify students’ sentiment on different issues. The way of using a questionnaire can also force students to answer the same question so that the data between different samples can have high consistency on aspects of content.

The ML tool used as predictor is the NN. For a predictor, one of the most important criteria is generalization performance, which means the prediction accuracy on fresh samples. However, the high dimension of samples will make itself too sparse to fill the sample space. In the training process of NN, the lack of samples can cause overfitting [6]. An overfitting NN fits the training set well but has poor prediction accuracy on fresh samples. As a result, a new way is needed to solve this problem. This paper will introduce a way that uses singular value decomposition (SVD) to reduce the dimension directly and add a closed loop based on genetic algorithm (GA) on the training process to relieve

overfitting. After obtaining a NN with good generalization performance, a method of calculating importance of each features is also proposed, which can help colleges combine macroscopic policies and microscopic guidance and strengthen the overall effectiveness.

Section 2 reviews the related work. Section 3 introduces the process of using SVD to pretreat the data set. Section 4 introduces the method of getting a predictor with good generalization performance, also the way of calculating features importance. Section 5 describes the details of the experiment and shows the results. Section 6 concludes our study and introduces future work.

2. Related Works

Early research studies on mining humans’ opinions have been done. Pang et al. [7] collected the review data from IMDb and used different tools of machine learning such as naive Bayes classification, maximum entropy classification, and support vector machines to classify audiences’ sentiment towards movies. Khan et al. [8] analyzed abundant text on Twitter that related to specific products and services and summarized the user’s overall views of those objects to help the producers and servers improve their works. Zhan et al. [9] designed an algorithm that not only mined opinion from customs reviews but also automatically pointed out the salient topics from these opinions, which can make the analysis more targeted. Zhou et al. [10] did the research to transfer customs’ reviews into answers of a questionnaire generated by the algorithm automatically and analyzed the collected data to point out what were the main points to improve user’s experience. Not only there are research studies focusing on objects, but also several others that try to focus on people. For example, Kosinski et al. [11] used “Facebook Likes” to predict a range of highly sensitive personal attributes and get high accuracy on some classification problems. Baik et al. [12] used buying behaviors to predict people’s score on four different personality traits and showed better precision when compared with previous studies. Besides the abovementioned research studies in different applications, some researchers also summarize the work in the whole field of public opinion mining. Pang and Lee [13] focused on improving the methods to address the new challenges raised by opinion mining. Tsytarau and Palpanas [14] tried to give a definition on opinion mining to clarify what is the basic work that should be done to mine public opinion. Ravi and Ravi [15] divided research studies into different levels and summarized the characteristics of each levels. These summaries provide researchers powerful tools to do opinion mining and give criteria to assess their work.

The method of using a questionnaire to collect data has been widely used in many situations when it is necessary to establish a person’s comprehensive personality profile. Topp et al. [16] reviewed 213 relevant articles to check the utility of a questionnaire named the WHO-5 Well-Being Index and confirmed its validity both in depression screening and outcome measuring in clinical trials. Garfinkel et al. [17] used a questionnaire to measure interoceptive sensibility,

which is an important dimension of one's interception. It could help explain cognitive, emotional, and clinical associations of interoceptive ability. Duckworth and Yeager [18] considered a self-report questionnaire is more efficient in studies of assessing internal psychological states like feelings of belonging when compared with other measures.

From previous research studies, it is clear that the method of using a questionnaire is good at collecting comprehensive data from a single person, and the data between different persons have high consistency on aspects. The collected data can be a good training material for human-focused opinion mining to learn the inner connection between students' behaviors and their state of mind. In this paper, the combination of the two methods overcomes the peculiarities and can make precise prediction on students' state of mind.

3. Data Collection and Pretreatment

This section will introduce what is the source of the data about college students' state of mind and describe the pretreatment method on data, including outlier detection and dimension reduction. Both of them are based on SVD.

3.1. Data Source. The data used in experiment come from a survey on students' state of mind that was conducted by Northwestern Polytechnical University in September, 2017. The students who had been surveyed were from different grades (including some masters and doctoral students). Under screening and checking, the total number of efficient sample data is 953.

The questionnaire consists of 30 questions, which are well designed to cover most aspects of students' daily life and their opinions. In terms of content, these questions can be divided as follows: (1) basic information: gender, grade, subject, and so on; (2) individual development: information of personal development since university entrance and future plan after graduation; (3) focus of attention: the focus of event happened recently; (4) mind identity: agreement on some policies and opinions; (5) school work evaluation: satisfaction with school work and direction of improvement. In terms of form, these questions can be divided into a single-choice question, multiple-choice question, scale question, and essay question.

Questions in different types need different primary pretreatments to get the original data set. Options in single-choice questions and multiple-choice questions are extended to independent variables, and the variable values were decided according to whether the options are selected or not; the answers of scale questions can be directly added into the data set; most of the questionnaires were left blank on essay questions so that they are ignored. After primary pretreatment, the sample vector dimension is extended to 160 dimensions. One of the variables is selected as sample label, and the rest are features of students. The sample label is given according to the students' evaluation on their own state of mind: the label 1 is positive, which means they do not need

to be guided; the label 0 means the students are not mature and need to be guided.

3.2. Meaning of SVD. SVD can be considered as the generalization of eigen decomposition from square matrix to matrix in any size [19]. In this case, the original data set is $S \in \mathbb{R}^{m \times n}$, which means there is m samples in the data set and each sample has n features. After the SVD process, there will be orthogonal matrixes $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ that present S as follows:

$$S = U\Sigma V^T. \quad (1)$$

In (1), Σ has the structure of $\Sigma = [\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n), 0]^T$ as $m > n$, where $\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ is a diagonal matrix and 0 is the zero matrix. σ_i is the singular values of S sorted in the descending order. If 0 is removed, the related vectors in U can be deleted so that $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ and $U \in \mathbb{R}^{m \times n}$.

An n dimension coordinate system can be established in the space of student samples whose axes relate to sample features, and every student samples can be represented by a point. The coordinate of sample is s_i , which is the row vectors of S . Then, the process of SVD can be considered as a coordinate transformation within the sample space, and each column vector v_i of V represents a base vector of the new coordinate system. The new base vectors can be given abstract meanings according to their relationship with original features. All the new base vectors are perpendicular to each other because V is an orthogonal matrix. Let $\hat{S} = U\Sigma$, so that

$$\hat{S} = SV. \quad (2)$$

From (2), it can be found that each row vector \hat{s}_i in \hat{S} represents the coordinate of a sample in the new coordinate system. Meanwhile, the singular values that relate to different base vectors represent the dispersion of samples on these directions. If the singular value is large, the samples' projections on its related base vector are widely distributed, which means there is abundant information stored.

3.3. Application of SVD in Outlier Detection. As larger singular value related to base vector which has a scattered distribution, it can be known that the bias on the base vector with small singular value will contribute more to a sample's deviation. As a result, the bias on base vector with small singular value should be given a high weight when calculating the total deviation of a sample. Before calculating sample's deviation, the singulars need to be sorted in the descending order as $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$. The calculation formula of weight is as follows:

$$w_j = \frac{\sum_{k=1}^j \sigma_k^2}{\sum_{k=1}^n \sigma_k^2}. \quad (3)$$

The bias of student sample i on new base vector v_j can be represented by Z -score. The calculation formula of Z -score is as follows:

$$z_{ij} = \frac{\widehat{s}_{ij} - \mu_j}{\sqrt{\sum_{k=1}^m (\widehat{s}_{ik} - \mu_k)^2}}, \quad (4)$$

where \widehat{s}_{ij} is the element of \widehat{S} and μ_j represents the mean of all elements in column vector $\widehat{s}_j = (\widehat{s}_{1j}, \widehat{s}_{2j}, \dots, \widehat{s}_{nj})$. The total deviation of the sample is calculated by the following equation:

$$d_i = \sum_{j=1}^n z_{ij} w_j. \quad (5)$$

After calculating deviations of all samples, a self-adapting threshold will be set. If a sample's deviation goes beyond the threshold, it will be deleted as outliers to make the data set more credible. A training set with high reliability will improve the generalization performance of the predictor.

3.4. Application of SVD in Dimension Reduction. It is found that larger singular value relates to more information, which means singular value can be used to help reduce the dimension of data set. The specific way to reduce dimension is to delete singulars with small values and its related vectors in U and V . Then, matrixes can be reconstructed as $\Sigma' = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$, $U' \in \mathbb{R}^{m \times k}$, and $V' \in \mathbb{R}^{n \times k}$. k is the number of reserved singulars, and formula (1) will be written as $= U' \Sigma' V'^T$.

However, even some new base vectors with small singulars might have high correlation with label, which means they can help increase the classification accuracy of the predictor. To protect them, the correlation between a base vector and sample label should be added in criterion. The importance score of a base vector is calculated by the following equation:

$$\text{score}_j = \sqrt{\sigma_j} \sum_{i=1}^n |c_i| v_{ij}^2, \quad (6)$$

where c_i is the correction between original features and label and v_{ij} is the element of V , which represent the relationship of original features and new base vectors.

The amount of information carried by a matrix can be measure by its Frobenius norm (F -norm). The F -norm of is calculated by the following equation:

$$\|S\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n s_{ij}^2} = \sqrt{\text{tr}(S^T S)} = \sqrt{\sum_{i=1}^n \sigma_i^2}, \quad (7)$$

where singular value σ_i is sorted by its score _{i} in the descending order. After base vectors with smaller scores have been deleted, the amount of remaining information can be represented by the F -norm of S' . And the percentage of the information reserved can be calculated by the following equation:

$$\text{percentage} = \frac{\|S'\|_F}{\|S\|_F} = \frac{\sqrt{\sum_{i=1}^k \sigma_i^2}}{\sqrt{\sum_{i=1}^n \sigma_i^2}}, \quad (8)$$

where k is the number of reserved base vectors.

The reduction on dimension of the sample space can prevent overfitting caused by sparsity of samples and strengthen the generalization performance of the predictor. Furthermore, because the noise carried by the data set is more likely to have smaller variance than the useful information, the dimension reduction can also weaken the impact of random noise on the data set.

4. Prediction on Students' State of Mind

This section will describe how the BP algorithm can be used in training NN for predicting students' state of mind. However, it is found using only BP algorithm will lead to overfitting, so a new algorithm which combines GA is proposed to relieve overfitting. After getting a NN with good generalization performance, a method of calculating importance of different features are also proposed.

4.1. BP-NN. BP algorithm is a common algorithm in ML. So, a NN trained by BP algorithm is established to predict the student's state of mind at first. After dimension reduction, the data of student samples can be represented by $\widehat{S}' \in \mathbb{R}^{m \times k}$. Here, m is not the total number of student samples, but the sample number after deleting outliers from the data set, and k is the number of remained new features of each sample student. Also, it should be $\widehat{S}' = U' \Sigma'$, but in fact, \widehat{S}' is Z -scored by (4) to fit the standard normal distribution on each features. This pretreatment will balance the learning rate of parameters in different nodes. Then, a data set $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$ is obtained, where $\mathbf{x}_i = \widehat{s}'_i$ is a row vector of \widehat{S}' and y_i is the label of the i th student sample.

The NN that is used to predict includes three layers. The input layer consists of k nodes for inputting the data vector \mathbf{x}_i . The output layer has only one node for outputting the prediction \widehat{y}_i of samples. The hidden layer's node number l is adjustable to fit the actual demand. I_i , H_h , and O , respectively, represent the i th input node, h th hidden node, and output node. The parameters of NN include connection weights ω_{ih} between I_i and H_h , connection weights ν_{ih} between H_h and O , thresholds γ_h of H_h , and threshold θ of O . The thresholds of nodes make NN become a nonlinear function, so that $f(\mathbf{x})$ is used as its equivalent function, and the output of NN is

$$\widehat{y}_i = f(\mathbf{x}_i). \quad (9)$$

The optimization goal of BP algorithm is usually the mean square error (MSE) between the output and label. The MSE can be calculated by the following equation:

$$E = \frac{1}{m} \sum_{i=1}^m (\widehat{y}_i - y_i)^2. \quad (10)$$

BP algorithm uses the strategy of adjusting parameters along the adverse direction of the gradient of E to decrease the error between prediction and real label. For example, the variation of ω_{ih} for each training round can be calculated by the following equation:

$$\Delta\omega_{ih} = -\mu \frac{\partial E}{\partial \omega_{ih}}, \quad (11)$$

where μ is the learning rate, which decides the speed of training.

Set the function between a student's features \mathbf{x}_i and his state of mind y_i as $F(\mathbf{x})$. The use of BP algorithm can help decrease the difference between $f(\mathbf{x})$ and $F(\mathbf{x})$ rapidly, so that the trained NN can be used as a predictor to make good prediction on student's state of mind.

4.2. Description and Analysis on Overfitting. However, BP algorithm did not work well in the primary experiment. To test the usefulness of the predictor, the data set D was divided into training set D_{train} and test set D_{test} randomly. It can be found from Figure 1 that the variation of the MSE of the NN's prediction on D_{train} and D_{test} shows difference.

As the number of training round increases, the MSE of NN on the training set approaches 0, which means the predictor fits the training set well. However, the MSE on the test set is still at a large value. This indicates that a well-trained NN may not have high prediction accuracy on fresh samples. $F1$ -measure (the harmonic mean of the recall and precision ratio) can represent the prediction accuracy of the predictor, and the mean $F1$ -measure on the training set is 0.97, while the mean $F1$ -measure on the test set is only 0.76. It means overfitting occurs.

Generally, the noise and unrelated features carried by the training set is considered as the reason of overfitting [20]. As the function between students' features and state of mind is set as $F(\mathbf{x})$, the influence on state of mind caused by noise and unrelated features can be defined as characteristic function $N(\mathbf{x})$, then the function of student samples in the training set is $F_{\text{train}}(\mathbf{x}) = F(\mathbf{x}) + N_{\text{train}}(\mathbf{x})$ and the function of student samples in the test set is $F_{\text{test}}(\mathbf{x}) = F(\mathbf{x}) + N_{\text{test}}(\mathbf{x})$.

All of the parameters in NN can be represented as $\mathbf{p} = (\omega_{11}, \dots, \omega_{kj}, \nu_1, \dots, \nu_l, \gamma_1, \dots, \gamma_l, \theta)$. Let MSE function $M(\mathbf{p})$ be the function from \mathbf{p} to the MSE between prediction and labels, and it can be known that the parameters of optimal NN is the global minimum point \mathbf{p}^g of $M(\mathbf{p})$. Due to the difference between $F_{\text{train}}(\mathbf{x})$ and $F_{\text{test}}(\mathbf{x})$, there will be difference between $M_{\text{train}}(\mathbf{p})$ and $M_{\text{test}}(\mathbf{p})$, thus difference between $\mathbf{p}_{\text{train}}^g$ and $\mathbf{p}_{\text{test}}^g$. If a NN selects $\mathbf{p}_{\text{train}}^g$ as its optimal parameter, it will fit $F_{\text{train}}(\mathbf{x})$ well, but the accuracy of its prediction on the test set may be not good. That is why overfitting occurs in BP-NN.

4.3. A Method of Relieving Overfitting. Although the experiment shows there is big difference between $\mathbf{p}_{\text{train}}^g$ and $\mathbf{p}_{\text{test}}^g$, $M_{\text{train}}(\mathbf{p})$ and $M_{\text{test}}(\mathbf{p})$ should be approximate. This can be verified by putting "prospect holes" on them. Putting "prospect holes" means to use same input \mathbf{p}_0 to predict different sample sets and compare the difference between $M_{\text{train}}(\mathbf{p}_0)$ and $M_{\text{test}}(\mathbf{p})$. After putting "prospect holes" randomly for 1000 times, the calculated mean percentage of MSE difference is 1.14%, which verifies the approximation. As a result, there can be several similar local minimum points \mathbf{p}^l in different $M(\mathbf{p})$ values. If the NN with parameters

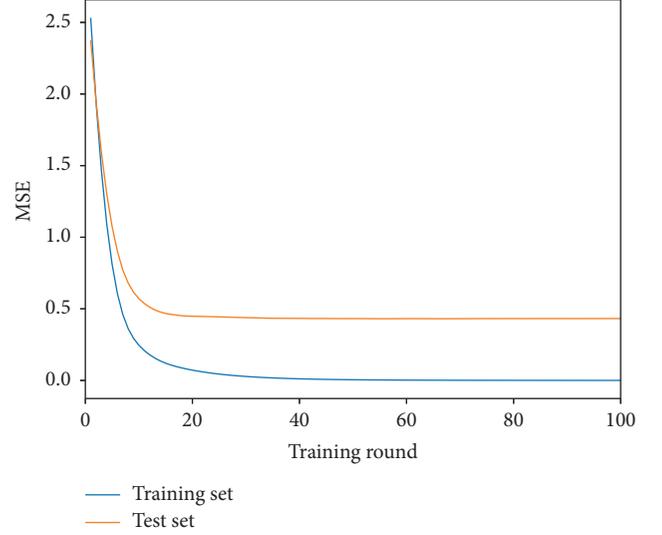


FIGURE 1: MSE on different data set.

of $\mathbf{p}_{\text{train}}^g$ does not fit the test set, then one of the $\mathbf{p}_{\text{train}}^l$ values which has high similarity with one of the $\mathbf{p}_{\text{test}}^l$ values can be used as the approximate optimal solution. Although the $M_{\text{train}}(\mathbf{p}_{\text{train}}^l)$ is slightly larger than $M_{\text{train}}(\mathbf{p}_{\text{train}}^g)$, it can fit both the training set and the test set well, which means to have good generalization performance.

In fact, the above method changes the criterion from only considering the MSE value to considering both the MSE values and the similarity with the test set. The optimization task of the MSE value can be handled by BP algorithm as usual. However, the similarity with the test set is difficult to quantify, but it can be indicated by $F1$ -measure of prediction on the test set.

To improve the similarity, an evolutionary algorithm is needed, so GA is introduced. It is known that different initial parameters of NN \mathbf{p}^i can make the network converge to different \mathbf{p}^l values when the training set is constant [21], and the training process can be represented as $\mathbf{p}^l = T(\mathbf{p}^i)$. So that \mathbf{p}^i can be regarded as an individual of population in GA, and population can be represented by $P = (\mathbf{p}_1^i, \mathbf{p}_2^i, \dots, \mathbf{p}_N^i)$ where N is the population size. After the NN with \mathbf{p}^i has been trained, the fitness of the individual will be calculated as the $F1$ -measure on the test set. After the operations of mate, mutate, and selection to generate, \mathbf{p}^i tends to fit both the training set and test set well.

However, using D_{test} to calculate individuals' fitness means D_{test} is also involved in the closed loop of algorithm, thus it loses the representation on fresh samples. To test whether the generalization performance of NN is improved, it is necessary to separate a set of samples before the algorithm to show the change of prediction accuracy on fresh samples. This sample set is called verification set D_{ver} .

After the primary experiment, if D_{test} which is used to calculate fitness is constant, the prediction accuracy of the test set will be improved greatly, but the prediction accuracy on verification set does not have a distinct change. This may be caused by the difference between $F_{\text{test}}(\mathbf{x})$ and $F(\mathbf{x})$. In order to make the algorithm effective, D_{test} is divided into

three parts as temporary test sets D_{temp} randomly, and a punishment is added if the network only performs well in one of the temporary test sets. Then, the fitness of individuals will be calculated as follows:

$$\text{fitness} = F1_{\text{mean}} - \sqrt{\frac{1}{3} \sum_{i=1}^3 (F1_i - F1_{\text{mean}})^2}. \quad (12)$$

The $F1_i$ in (12) is the $F1$ -measure on the i th D_{temp} , and $F1_{\text{mean}}$ is the mean of $F1_i$. After modification of the algorithm, each individual faces different D_{temp} values when calculating fitness. This method will dilute $N_{\text{test}}(\mathbf{x})$, so that the evolutionary direction is to fit $F(\mathbf{x})$ rather than to fit $F_{\text{test}}(\mathbf{x})$. It can make sure that the NN is going to have better generalization performance.

It has been found that overfitting when predicting students' state of mind is caused by difference between $F_{\text{train}}(\mathbf{x})$ and $F(\mathbf{x})$, so it will be of benefit to use the above algorithm. Just as the process represented in Figure 2, the whole data set is divided into D_{train} , D_{test} , and D_{ver} at first. Then, the initial P is generated randomly, and all of the NNs with \mathbf{p}^i will be trained by the same training set. To keep genetic advantage of the individual with high fitness, elite strategy is used when generating the next population. This strategy produces offsprings by mating and mutating before selecting individuals in the next population, and the offspring has the same size of current population which is N . All of the individuals in the current population and offsprings are sorted by their fitness in the descending order, and the first N individuals are selected as the next population. When the population reaches the largest generation, the individual that relates to the NN who has the largest prediction accuracy on D_{ver} will be chosen as optimal solution \mathbf{p}^o . Finally, a new network whose initial parameter is \mathbf{p}^o will be trained by the whole data set to get the predictor that can predict students' state of mind with high accuracy.

4.4. Feature Importance Calculation. BP algorithm combined with GA helps obtain a predictor on students' state of mind with high accuracy, but the predictor is a "black box" which means the mechanism of making prediction is still unknown. Though NN is known as an unilluminated method, there are still some ways to evaluate the importance of different features [22]. After a NN is trained by D , it has nearly perfect prediction accuracy on D . But if one feature is sheltered (replace the mean of this feature, which is 0 after z -scored) in each student sample, there would likely be a recession on accuracy [23]. Then, the importance of features in D can be calculated by the following equation:

$$im_i = F1'_i - F1_0, \quad (13)$$

where $F1_0$ is the accuracy before sheltering and $F1'_i$ is the accuracy after i th has been sheltered. The features in D are abstraction of original features, which means the importance of original features can also be calculated according to V' which represents the relationship between new features and original features by the following equation:

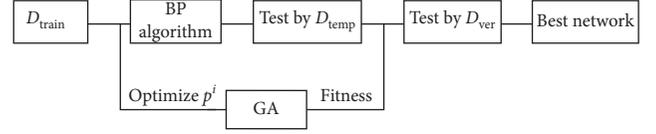


FIGURE 2: Structure of the algorithm.

$$im'_i = \sum_j = 1^k im_j v_{ij}. \quad (14)$$

The importance analysis on single NN lacks credibility, so the total importance of original features is accumulated after analyzing 500 NNs. The importance can help colleges know what is important in guiding students' state of mind.

5. Experiment Results and Discussion

This section will show the experiment results which can support the hypotheses that have been proposed above. They can also show the actual effectiveness of this new method on predicting students' state of mind.

5.1. Implementation Details. The first step of experiment is the pretreatment of data. The raw data collected are already numbered according to the order of answers or whether an answer has been ticked, and the data set expansion strategy described in Section 3.1 is used to make it regular. Then, the whole data set is regard as a matrix, and SVD is performed. The attained matrixes after SVD are used in the process of outlier detection and dimension reduction which are described in Sections 3.3 and 3.4. After pretreatment, each sample is represented by a vector in the new coordinate system and a label of state of mind. Then, the samples are used to start the operation of the algorithm described in Section 4.3. The algorithm will train many NNs, and the NN with best generalization performance will be selected to make precise prediction on students' state of mind. The rest of the trained NNs also show part of inner connection between students' behaviors and their state of mind, so all of the trained NNs are used to calculate importance of different features by the method described in Section 4.4.

5.2. Experiment on Data Pretreatment. The pretreatment of data set includes outlier detection and dimension reduction. Both of them have been detailed in Section 3. After calculating deviations of all samples, the distribution of deviations is represented in Figure 3. It can be found that the distribution of sample's deviation roughly conforms to the normal distribution. The normal distribution which is shown in Figure 3 is obtained by fitting the original distribution approximately. The mean of normal distribution is $\mu = 105.18$, and the standard deviation is $\sigma = 13.97$. Then the self-adaption threshold is calculated by $\text{threshold} = \mu + 2\sigma = 133.12$, and the number of outliers is 27.

After checking the content of deleted outliers, many of them are found with contradictions in context. For example, a student said his counselor is the one who gives the best help in development of his state of mind, but he also said he was

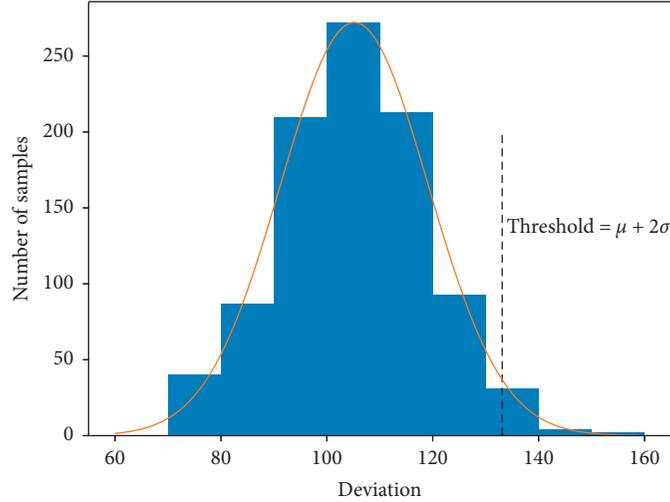


FIGURE 3: The distribution of samples' deviation.

dissatisfied with counselors' work. This might be his real thought, but will confuse the predictor, so he is removed from the sample set. Other outliers can make mistakes while answering questions, such as making multiple choices on single-choice question. It will also change the potential importance of a variable so that they should be removed. The experiment result on outlier detection shows that the algorithm is effective and reasonable. The application of outlier detection will purify the data set and help improve the generalization performance of the predictor.

Figure 4 shows the change in the number of reserved inputs with the increase in percentage, which represents the percentage of reserved information calculated by F -norm. It can be found that the number of inputs have a sharp increase when the percentage is large. It confirms that there are many inputs related to small singular values and have a low correlation with label that should be deleted to reduce the dimension of sample space. In experiment, 90% information is reserved, and the number of inputs changes from 159 to 71.

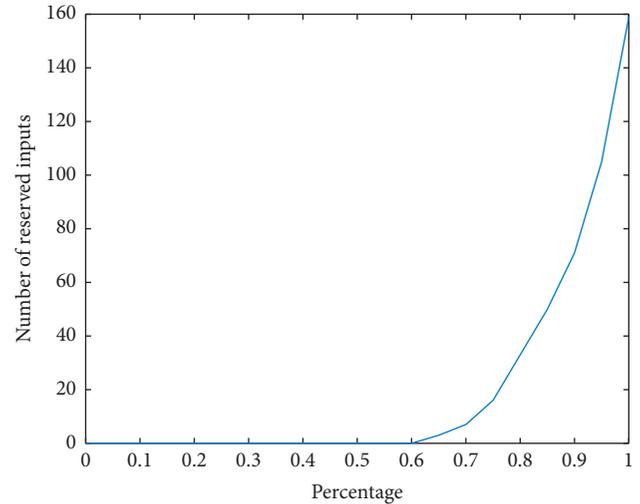


FIGURE 4: Number of reserved inputs in different percentages.

5.3. Experiment Results on Overfitting Relief. The overfitting relief experiment is designed according to the algorithm mentioned in Section 4. Abnormal samples have been previously excluded by the outlier detection based on SVD, and 926 samples are used. Under the principle of randomness, 70% samples are selected as D_{train} , 20% samples are selected as D_{test} , and the rest 10% are used as D_{ver} . The number of individuals in a population is set to 10, and the maximum number of generations is set to 50. The population crossover rate is 0.7, and the mutation rate is 0.3. Figures 5(a) and 5(b) show the change on population's mean $F1$ -measure on different data sets when using constant D_{test} to calculate individuals' fitness, and Figures 6(a) and 6(b) show the changes when using D_{temp} to calculate fitness.

Figures 5(a) and 5(b) show the different tendency of population's mean $F1$ -measure on D_{test} and D_{ver} . It can be found the prediction accuracy on D_{test} has an obvious increase, but the prediction accuracy on D_{ver} decreases slightly, which means population tends to fit $F_{\text{test}}(\mathbf{x})$ when using D_{test}

to calculate fitness. If D_{temp} s are used, it can be found in Figure 6(a) that the prediction accuracy on D_{test} does not reach the same increase shown in Figure 5(a). But on D_{ver} , the prediction accuracy increases obviously. This tells the hypothesis in Section 4 is reliable.

In Figure 6(b), it can be observed that the mean $F1$ -measure of the initial population on the verification set is 0.7662. After the evolution, the mean $F1$ -measure of the last generation reaches 0.8080. When applying the network as a predictor in real engineering, the one with best accuracy from the current population will be chosen. The data show that the biggest $F1$ -measure is 0.8315, which represents high prediction accuracy. With reliable prediction of students' state of mind, colleges can make effective guidance to help students get rid of bad influences of information explosion.

5.4. Experiment on Feature Importance. To evaluate the importance of different students' features, 500 NNs are trained by the whole data set D . The calculated importance of

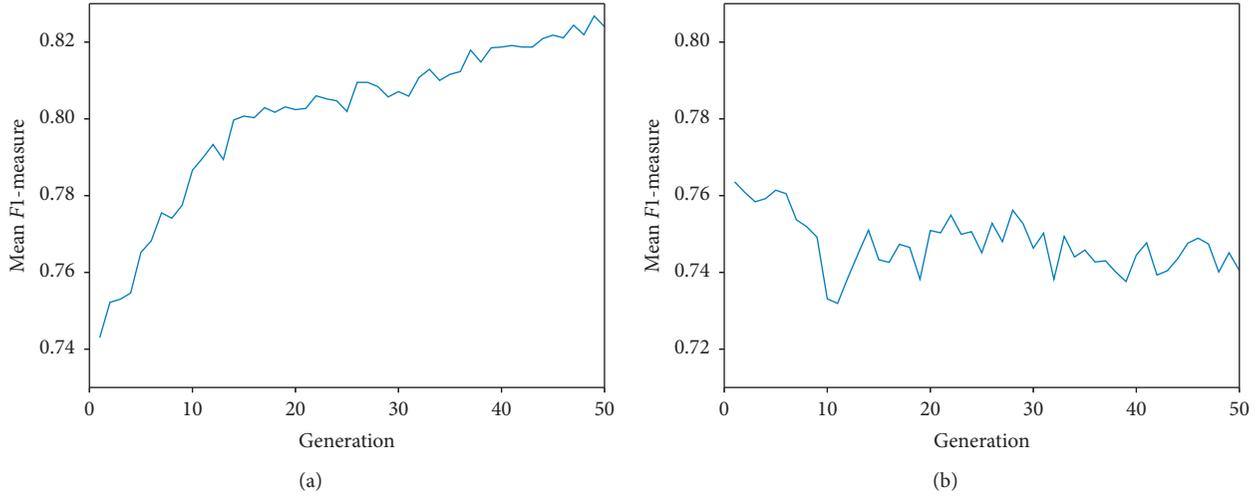


FIGURE 5: Using constant test set to calculate fitness. (a) Test set. (b) Verification set.

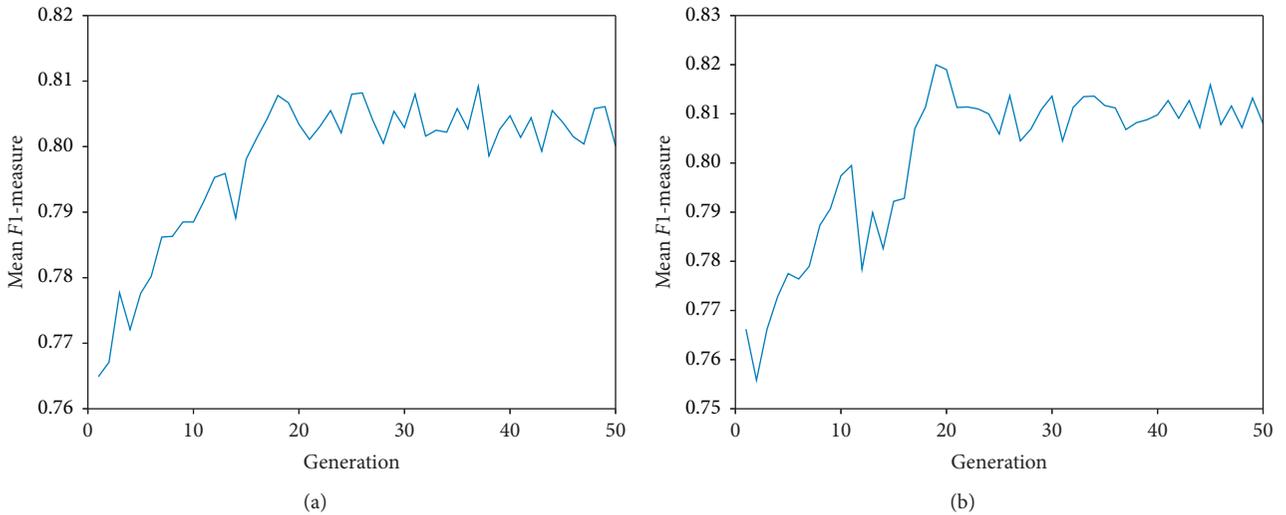


FIGURE 6: Using temporary test set to calculate fitness. (a) Temporary test set. (b) Verification set.

original features are partly shown in Table 1. The features are sorted by their absolute importance and only top 10 are listed in Table 1. It can be found some features positively relate to students' state of mind while others negatively. Also, some features show a potential causal relationship with state of mind such as "focus on politics news" and "benefit more from ideology education," while others may not such as "use QQ often" and "benefit more from club activities" though they do have correlation. It can also be noticed that the features classified as "individual development" have the best power in indicating students' state of mind.

With analysis on feature importance, colleges will be aware of what are the key points of guidance on students' state of mind. For example, they can help students interest in politics news, strengthen the relationship between students and their tutors, and enrich social practice. These macroscopic policies can combine with the microscopic guidance on students' state of mind and enhance overall effectiveness.

6. Conclusion

The main purpose of this paper is to change the traditional method on mining humans' opinion to make it effective when predicting students' state of mind. This changed method requires more data aspects of samples, and using a questionnaire is a good choice to get comprehensive data about students. However, the expansion of sample space's dimension makes samples sparser and causes overfitting while training NN. To solve this problem, SVD is used to reduce the dimension of sample space directly, and a closed loop based on GA is added to help NN have better prediction accuracy on fresh samples. The result of the experiment shows that the new algorithm works well and the predictor obtained has good generalization performance. Also a simple method of calculating features' importance is proposed, which can help colleges make policies.

The new method lets the predictor make reliable predictions on students' state of mind. With these predictions,

TABLE 1: Feature importance.

Rank	Importance	Description	Classification
1	1.5069	Focus on politics news	Individual development
2	0.7620	Benefit more from tutor	Individual development
3	0.7487	Focus on the broadcasting of <i>The truth of 731 troop</i> in NHK	Focus of attention
4	0.7358	Think addiction to smartphone is common	Individual development
5	-0.7055	Focus on relaxation news	Individual development
6	-0.6830	Benefit more from club activities	Individual development
7	-0.6364	Use QQ often	Individual development
8	0.6281	Satisfy on school's politics education	School work evaluation
9	0.5388	Benefit more from ideology education	Individual development
10	0.5348	Benefit more from social practice	Individual development

colleges can apply guidance associated with students' personal experience which will make it more genial and effective. Furthermore, the macroscopic policies made according to feature importance can supplement microscopic guidance to have better effectiveness.

For further research, a questionnaire used to collect data will be redesigned. The aim of questions in the questionnaire should be more covert to make sure to collect real information, and the content of questions should be more various especially in "individual development" to collect more data that might be necessary. Also, the classification problem on students' state of mind will be changed into quantization problem to get a student's certain score on different aspects of state of mind. The method of calculating feature importance will be improved too. These future studies will be able to further strengthen the effect of guidance on students' state of mind.

Data Availability

The students' state of mind data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (3102018jcc041 and 3102018jcc028) and the 2011 Collaborative Center Innovation Project under Grant QBXT-17-7.

References

- [1] N. A. Cheever, L. D. Rosen, L. M. Carrier, and A. Chavez, "Out of sight is not out of mind: the impact of restricting wireless mobile device use on anxiety levels among low, moderate and high users," *Computers in Human Behavior*, vol. 37, pp. 290–297, 2014.
- [2] M. Samaha and N. S. Hawi, "Relationships among smartphone addiction, stress, academic performance, and satisfaction with life," *Computers in Human Behavior*, vol. 57, pp. 321–325, 2016.
- [3] L. Zhao, J. Yin, and Y. Song, "An exploration of rumor combating behavior on social media in the context of social crises," *Computers in Human Behavior*, vol. 58, pp. 25–36, 2016.
- [4] Z. Demirtas, "The relationship between school culture and student achievement," *Egitim Ve Bilim-Education and Science*, vol. 35, no. 158, pp. 3–13, 2010.
- [5] L. Yu and S. Ding, "Broad learning based optimization and prediction of questionnaire survey: application to mind status of college students," *Complexity*, vol. 2018, Article ID 5736030, 9 pages, 2018.
- [6] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent NNs from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [7] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing—EMNLP '02*, J. Hajic and Y. Matsumoto, Eds., vol. 10, pp. 79–86, Association for Computational Linguistics, Stroudsburg, PA, USA, July 2002.
- [8] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," *Decision Support Systems*, vol. 57, pp. 245–257, 2014.
- [9] J. Zhan, H. T. Loh, and Y. Liu, "Gather customer concerns from online product reviews—a text summarization approach," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2107–2115, 2009.
- [10] Y. Zhou, J. Chen, and Y. Kuo, "Fairness resource allocation for parallel multi-radio access in cognitive multi-cell," *Wireless Personal Communications*, vol. 88, no. 3, pp. 587–602, 2016.
- [11] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 15, pp. 5802–5805, 2013.
- [12] J. Baik, K. Lee, S. Lee, Y. Kim, and J. Choi, "Predicting personality traits related to consumer behavior using SNS analysis," *New Review of Hypermedia and Multimedia*, vol. 22, no. 3, pp. 189–206, 2016.
- [13] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 1, no. 2, pp. 1–135, 2008.
- [14] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," *Data Mining and Knowledge Discovery*, vol. 24, no. 3, pp. 478–514, 2012.
- [15] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14–46, 2015.

- [16] C. W. Topp, S. D. Østergaard, S. Sondergaard, and P. Bech, "The WHO-5 well-being Index: a systematic review of the literature," *Psychotherapy and Psychosomatics*, vol. 84, no. 3, pp. 167–176, 2015.
- [17] S. N. Garfinkel, A. K. Seth, A. B. Barrett, K. Suzuki, and H. D. Critchley, "Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness," *Biological Psychology*, vol. 104, pp. 65–74, 2015.
- [18] A. L. Duckworth and D. S. Yeager, "Measurement matters: assessing personal qualities other than cognitive ability for educational purposes," *Educational Researcher*, vol. 44, no. 4, pp. 237–251, 2015.
- [19] G. H. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *Journal of the Society for Industrial and Applied Mathematics*, vol. 2, no. 2, pp. 205–224, 1965.
- [20] D. M. Hawkins, "The problem of overfitting," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 1, pp. 1–12, 2004.
- [21] C. Ren, N. An, J. Z. Wang, L. Li, B. Hu, and D. Shang, "Optimal parameters selection for BP NN based on particle swarm optimization: a case study of wind speed forecasting," *Knowledge-Based Systems*, vol. 56, pp. 226–239, 2013.
- [22] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2018.
- [23] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2660–2673, 2017.