Psychological Crisis Prediction of Students Based on Network Behavior by Big Data Mining

https://doi.org/10.3991/ijet.v18i12.41091

Zhiping Jia(✉)
Department of Student Affairs, Hebei Chemical & Pharmaceutical College, Shijiazhuang, China
jzp123004@163.com

Abstract—Signs of psychological crisis can be found in time by analyzing the network behavior data of college students, thus providing a basis for early warning and intervention. However, existing methods may not only have shortcomings in handling dynamic data and updating models, but also rely too much on network behavior data and overlook other factors possibly affecting the psychological crisis of college students. In order to overcome these shortcomings, this paper aimed to study the psychological crisis prediction of college students based on big data mining of network behavior. Network behavior interactive prediction was defined to determine the objective function of the constructed model. Interactive prediction model framework was presented and the working principle of the model was explained. Finally, various early warning indexes, which needed to be comprehensively considered in the psychological crisis early warning model of college students, were given, and the combination of principal component analysis (PCA) and support vector machine (SVM) was applied to the construction of the early warning model, thus improving its prediction effects, generalization ability and interpretability, and reducing the overfitting risk and the difficulty of processing high-dimensional data. The experimental results verified that the constructed model was effective.

Keywords—network behavior analysis, data mining, psychological health of college students, psychological crisis prediction

1 Introduction

In recent years, psychological problems of college students have increasingly attracted attention from various sectors of society [1–3]. In modern society, college students are faced with enormous pressure in their studies, daily life, and employment, which easily leads to their psychological problems [4–9]. At the same time, network behavior is an important way to reflect the psychological status of college students, because the rapid development of network technology has made their behavior increasingly diverse in the network. Therefore, the psychological crisis prediction of college students based on big data mining of network behavior is of important theoretical and practical significance. Network behavior data analysis of college students identifies psychological crisis signs in time, and provides a basis for early warning and intervention [10–18], thus helping reduce their psychological crisis incidence and improving their mental health level. The research results can provide data support and theoretical basis for mental health education in universities, which enables the universities to adjust their mental health education strategies accordingly and improve the pertinence and effectiveness of mental health education.

Although mental health education of students attracts more and more attention, there has been neither extensive research on the stressors of psychological crisis, nor targeted research on the psychological crisis early warning of college students with varying degrees of psychological crisis. To solve this problem, Jia [19] explored the psychological crisis early warning of college students and sports intervention using artificial neural network. First, the study determined important evaluation indexes for the early warning, and processed the index data using partial least squares, based on the adverse reactions and performance of college students in physiological, cognitive, emotional, and behavioral aspects. Second, a psychological crisis early warning model was established based on the optimized neural network. The study improved the harmony search algorithm based on differential evolution algorithm, and used it to optimize the back propagation neural network. The experimental results verified that the model was feasible and effective. Liu et al. [20] explored the application of big data technology in current psychological management system by investigating psychological crisis screening indexes. Data mining technology was used to dynamically manage psychological early warning data, and monitor the psychology of high risk group in real time, thus improving the accuracy and effectiveness in early identification and warning of psychological crisis of students. By combining qualitative and quantitative analysis, a series of studies were conducted on three typical network consensus types, namely, Internet rumors, network consensus of college students, and public health emergencies, in terms of transmission mechanism, early warning, decision-making mechanism, and evolution mechanism of network consensus. Based on quantitative statistical analysis,

Although the psychological crisis prediction method of college students based on big data mining of network behavior has achieved significant results in many aspects, it still has some shortcomings. Psychological crisis identification usually requires the evaluation of professional psychologists, and this evaluation process is subjective and complex. The psychological state and network behavior of college students have strong dynamism and timeliness. Therefore, it is necessary to continuously update the model in order to adapt to the constantly changing data environment. However, existing methods may have shortcomings in handling dynamic data and updating models. In addition, these methods may overly rely on network behavior data and overlook other factors, which may affect the psychological crisis of college students, such as family background and individual personality, thus may lead to limitations and one-sidedness in the prediction results. In order to overcome these shortcomings, this paper studied the psychological crisis prediction of college students based on big data mining of network behavior. After defining the network behavior interactive prediction of college students, Chapter 2 determined the objective function of the constructed model. Chapter 3 presented the interactive prediction model framework and explained the working principle of the model. Chapter 4 gave a variety of early warning indexes to be comprehensively considered in the early warning model, and applied the combined PCA with SVM for modeling, thus improving the prediction effects, generalization ability and interpretability of the model, and reducing the overfitting risk and the difficulty of processing high-dimensional data. The experimental results verified the effectiveness of the constructed model.

2 Definition of network behavior interactive prediction of college students

Based on big data mining of network behavior, this study divided the psychological crisis prediction process of college students into two parts: network behavior interactive prediction, and construction of psychological crisis early warning model.

In terms of network behavior interactive prediction, network behavior data of college students was first collected from various network platforms, such as social media, forums, blogs, and so on, including posted content, comments, "liking", and reposting. Key features of the network behavior were extracted, such as posting frequency, keywords, emotional tendencies, etc. Then recurrent neural network (RNN) was used to model the network behavior data of college students, because RNN processed sequence data and captured the network behavior variation features in time dimension.

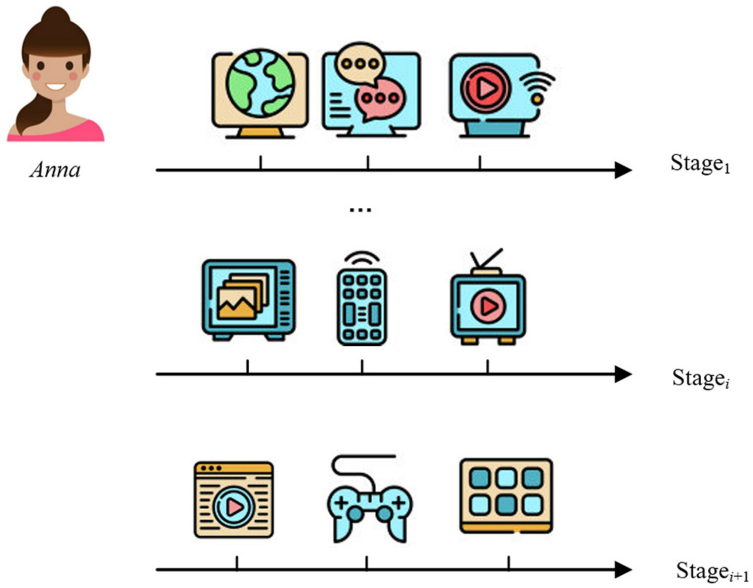

Fig. 1. Schematic diagram of student network behavior interactive sequence

Figure 1 shows the schematic diagram of student network behavior interactive sequence. Network behavior big data used for predicting the psychological crisis of college students mainly included the following aspects: 1) Text data: text content from social media, forums, blogs, chat transcripts, etc., such as posts, comments, private messages, etc., which may contain keywords and emotional tendencies expressing the psychological status of college students. 2) Interactive behavior data of college students on network platforms, such as "liking", reposting, following, favoriting, etc., which reflected their social network status and interaction preferences. 3) Time data: network behavior time of college students, such as time of posting and commenting, etc., which helped analyze the patterns and periodicity of their network behavior. 4) Login and online time data: frequency and online time of college students logging into network platforms, which reflected their activity level and time distribution in the cyberspace. 5) Personal information data: personal information of college students on network platforms, such as age, gender, university, major, etc., which helped analyze the relationship between their psychological status and personal background. 6) Multimedia data: multimedia content, such as images and videos posted by college students on network platforms, which provided intuitive information on their psychological status. 7) Behavior track data, such as browsing and search records of college students on network platforms, which helped understand their interests, hobbies, and focus of attention.

Based on the above thinking and the network behavior big data content, used for psychological crisis prediction of college students, this study defined the interactive prediction between students and interactive projects on network platforms. Let $v_p \in R^m \forall v \in V$ be the student embedding, $i_p \in R^m \forall i \in J$ be the interactive project embedding, V be the student collection, J be the interactive project collection, $R(v, i, p, g) \in R$ be an ordered interactive sequence of students and interactive projects, v be one student in V, i be one interactive project in J, p be the timestamp of R, and g be the corresponding feature vector of each interaction between students and interactive projects on network platforms. Figure 2 shows the network behavior interaction with timestamps.

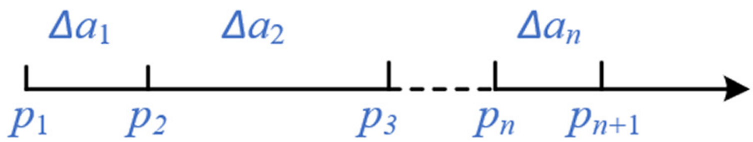

Fig. 2. Network behavior interaction with timestamps

Given the historical interaction SR of a set of students and interactive projects, the network behavior interaction prediction of college students aimed to learn the future student embedding $v_{p+\Delta}$ and the future interactive project embedding $i_{p+\Delta}$, and predict the interactive behavior of student v with which interactive project at $p + \Delta$ time in the future.

Let $Q_1, \ldots, Q_4$ be the trainable parameters, χ be the bias vector, v^- and i^- be the static embeddings of students and interactive projects, $v_{p+\Delta}$ be the output of the shift embedded module, and $i_{p+\Delta-1}$ be the embedding of interactive projects before the $p + \Delta$ time, then the prediction function of interactive project embedding in the prediction model constructed in this study was as follows:

$$j_{p+\Delta} = Q_1 v_{p+\Delta} + Q_2 \overline{v} + Q_3 i_{p+\Delta-1} + Q_4 \overline{i} + \chi \tag{1}$$

In order to train the parameters Q_1, ..., Q_4 in the constructed model, this study minimized the *L*2 difference between the predicted and the actual interactive project embedding $\tilde{j}_p$ and j_i, when students interacted with interactive projects on network platforms. Let μ_V and μ_p be the scaling parameters, then specifically the constructed model aimed to minimize the loss function shown in the following formula:

$$\min \sum_{(v,j,p,g)\in R} \left\| \tilde{j}_p - j_p \right\|_2 + \mu_V \left\| v_p - v_{p-1} \right\|_2 + \mu_I \left\| i_p - i_{p-1} \right\|_2 \qquad (2)$$

3 Interactive prediction between students and interactive projects based on RNN

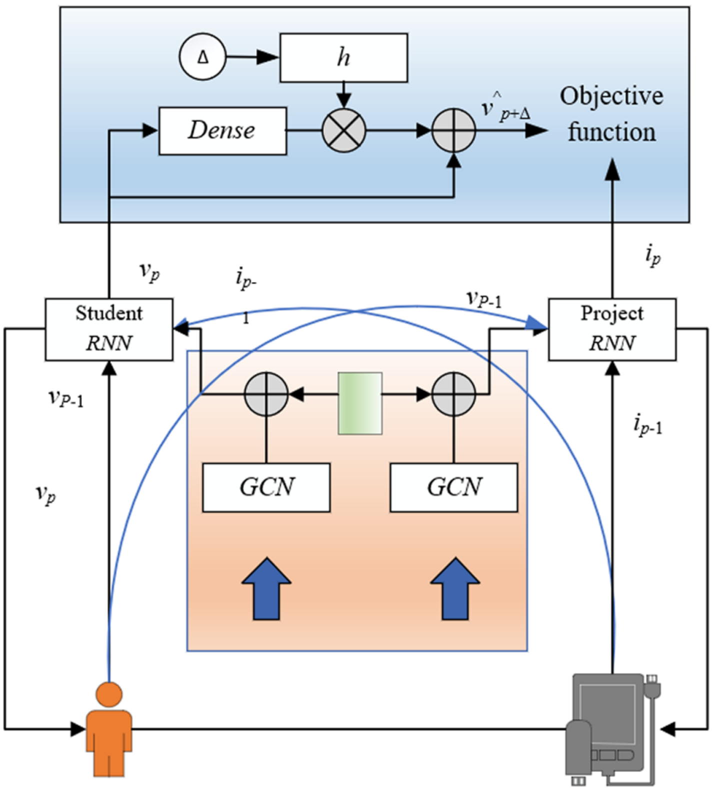

Fig. 3. Interactive prediction model framework of network behavior of college students

The RNN used for network behavior interactive prediction of college students consisted of three modules, namely, graph convolutional network (GCN) module, RNN module, and shift embedded module. First, the relationship between students and interactive projects was captured using the GCN module, and their representations were obtained. Then the embedding of students and interactive projects was dynamically updated using the RNN module in order to capture their changing characteristics over time. Finally, the shift embedded module was used to predict the behavior track of students at a certain point in the future, thus achieving network behavior interactive prediction of college students. This method integrated student-interactive project relationships, dynamic features, and temporal information, which helped improve the prediction accuracy and generalization ability. Figure 3 shows the interactive prediction model framework of network behavior.

The GCN module was used to capture auxiliary information between students and interactive projects. By learning on the relationship graph between students and interactive projects, the GCN module obtained their representations, which were constructed based on the original interactive graph between students and interaction projects according to second-order similarity. The second-order similarity represented the similarity between students or between interactive projects, which helped explore potential connections. Let $X_v \in R^{|V|\times|V|}$ and $X_i \in R^{|J|\times|J|}$ be the adjacency matrices of the second-order similarity graph between students and between interactive projects; element X_v^{ij} in the *i*-th row and *j*-th column in X_v be the number of interactive projects that have interacted with both student *i* and student *j* simultaneously in the entire interactive sequence collection of students and interactive projects ξ; $|V|$ and $|J|$ be the number of students and interactive projects in student collection V and interactive project collection J, respectively; $O_v \in R^{|V|\times cv}$ and $O_i \in R^{|V|\times ci}$ be the static feature matrices of students and interactive projects, which were constructed based on their corresponding static features, with c_v and c_i as feature dimensions.

O_v and O_i were updated by aggregating the representations of adjacent entities using the feedforward layer. Let F_v^k and F_i^k be the hidden representation matrices of students and interaction projects in *k*-th layer; $F_v^0 = O_v$ and $F_i^k = O_i$ be the initialized representation matrices in 0 layer; X_v and X_i be the adjacency matrices of student *v* and interactive project *i*; C_v and C_i be the diagonal matrices of X_v and X_i; Q_v^k and Q_i^k be the trainable weight matrices in *k*-th layer; ε be a nonlinear activation function; and K be the total number of layers. The feedforward propagation expressions between layers were given as follows:

$$\begin{aligned} F_v^{k+1} &= \varepsilon(C_v^{-1/2} X_v C_v^{-1/2} F_v^l Q_v^l), k = 0,1,...,K-1 \\ F_i^{k+1} &= \varepsilon(C_i^{-1/2} X_i C_i^{-1/2} F_i^k Q_i^k), k = 0,1,...,K-1 \end{aligned} \tag{3}$$

Let F_v^K and F_i^K be the final output of the GCN module; g be the interactive feature vector from students to interactive projects; e_v and e_i be their combinations, which were used as auxiliary features for learning dynamic student and interactive project embeddings.

Two RNN modules were used to dynamically update the embeddings of students and interactive projects, respectively. The two modules projected the embeddings of students and interactive projects into the same embedded space through mutual learning. The RNNs captured the changing characteristics of student and interactive project behavior over time, because they were able to process temporal data. Based on this idea, let v_p and i_p be the dynamic embedding of student u and interactive project i at p time; e_v and e_i be the auxiliary features learned from the GCN module; σ be a sigmoid function; $Q_1^v, \ldots, Q_2^v$ be the parameter matrix of student RNN; $Q_1^i, \ldots, Q_3^v$ be the parameter matrix of interactive project RNN. The embedding iteration update formulas of students and interactive projects were given as follows:

$$\begin{aligned} v_p &= \varepsilon(Q_1^v v_{t-1} + Q_2^v i_{p-1} + Q_3^v e_v) \\ i_p &= \varepsilon(Q_1^i i_{p-1} + Q_2^i i_{p-1} + Q_3^v e_i) \end{aligned} \tag{4}$$

Shift embedded module was used to predict the student embedding at a certain point in the future, which predicted the student embedding at $p + \Delta$ time in the future, based on the student embedding at current time, thus predicting the student behavior track in the future. Let Δ be the time interval from the last interaction between students and interactive projects, Q_R be the parameter matrix of the linear layer, χ be the bias, $h(\Delta) = Q_t \cdot log\ (o + \Delta)$ be the function used to convert c into a context vector about time, Q_t be the trainable parameter, and $\hat{v}_{p+\Delta}$ be the predicted student embedding at $p + \Delta$ time. After adjusting the internal memory v_p^R, which was obtained using linear layer based on time interval information, this study obtained $\tilde{v}_p^R$, which formed shifted student embedding after combining with v_p^R. The detailed representations of the shift embedded module were given as follows:

$$\begin{aligned} v_p^R &= \varepsilon(Q_R v_p + \chi) \\ \tilde{v}_p^R &= v_p^R * h(\Delta) \\ \hat{v}_{p+\Delta} &= v_p + \tilde{v}_p^R \end{aligned} \tag{5}$$

4 Construction of psychological crisis early warning model of college students

Multiple early warning indexes needed to be comprehensively considered in the psychological crisis early warning model of college students. Combined with the network behavior interactive prediction results, the following eight early warning indexes were selected for the constructed model: 1) network behavior index, including posting frequency, keyword usage, emotional tendencies, interactive behavior, such as "liking", comments, reposting, etc., which reflected the activity level and psychological status of college students in the cyberspace. 2) Academic performance index, such as average scores, number of failed courses, etc., which reflected the learning status of college students, because academic stress may be related to their psychological crisis. 3) Social relationship index, such as size of social circle, social activity participation, etc., which reflected the social status of college students, because loneliness and social support may have an impact on their mental health. 4) Living habit index, such as regular routines, healthy diet, etc., which reflected the quality of life of college students, because bad living habits may increase their psychological stress. 5) Family background index, such as family economic status, family relationships, etc., which reflected the family environment of college students, because family factors may affect their mental health. 6) Individual personality index, such as introversion, extroversion, neuroticism, etc., which reflected the personality characteristics of college students, because certain personalities may be more likely to lead to their psychological problems. 7) Psychological test index, such as anxiety, depression, stress and other psychological test scores, which directly reflected the psychological status of college students and provided important basis for the early warning model. 8) Historical psychological crisis event index, such as past psychological counseling records, psychological crisis events, etc., which helped identify college students at risk of psychological problems.

In terms of model construction, this study first integrated the predicted network behavior interaction data of college students with existing psychological crisis label data, thus forming a complete training set. PCA method was used to reduce the dimensions of the integrated data, extract the main features, and reduce the computational complexity. SVM was used to construct a psychological crisis early warning model of college students, based on the dimension-reduced data, because SVM had strong classification ability and effectively distinguished college students with different psychological states. The trained SVM model was used for psychological crisis early warning. Timely psychological intervention and support were provided for college students according to the warning results. Application of combined PCA with SVM to the model construction not only improved the prediction effects, generalization ability and interpretability of the model, but also reduced the overfitting risk and the difficulty of processing high-dimensional data.

It was assumed that there were n rows of psychological crisis label data with m feature dimensions, and a_{ij} represented the j-th dimension attribute of the i-th row data, then A was an $n \times m$ matrix, which was constructed based on a_{ij}. Let V be the matrix composed of eigenvectors of the covariance matrix D corresponding to A, which was the matrix of $m \times l$. Data matrix of $C = AV$ was obtained through dimensionality reduction of A, where C is the matrix of $n \times l$. Let g_j ($j = 1, 2, \ldots, n, n \leq m$) be the linear combination principal component, then there were:

$$\begin{cases} g_1 = \phi_{11}a_1 + \phi_{12}a_2 + \cdots + \phi_{1m}a_m \\ g_2 = \phi_{21}a_1 + \phi_{22}a_2 + \cdots + \phi_{2m}a_m \\ \vdots \\ g_n = \phi_{n1}a_1 + \phi_{n2}a_2 + \cdots + \phi_{nm}a_m \end{cases} \tag{6}$$

In order to minimize the information loss from the principal components, coefficient φ_{ij} should be selected appropriately to ensure that g_j met:

(1) No correlation between g_i and g_j ($i \neq j$);
(2) g_1 had the largest variance among various linear combinations of $a_1, a_2, \ldots\ldots, a_m$. And g_2 was not correlated to g_1, and had the largest variance among various linear combinations of $a_1, a_2, \ldots\ldots, a_m$. By analogy, g_2 was not correlated to $g_1, g_2, \ldots\ldots, g_{n-1}$, and had the largest variance among various linear combinations of $a_1, a_2, \ldots\ldots, a_m$.

Therefore, $g_1, g_2, \ldots\ldots, g_n$ were considered as the first to n-th principal components of a_i($i = 1, 2, \ldots\ldots, m$).

Let $\{(a_1, b_1), (a_2, b_2), ..., (a_m, b_m)\}$ be the initial training set, $a_i \in R^m$ and $b \in \{-1,1\}$. Based on the idea of kernel method, kernel function $L(a_i \cdot a_j)$ satisfying Mercer theorem was found in the input space. For ease of calculation, dot product operation $\psi(a_i)\psi(a_j)$ in high-dimensional feature space was replaced with the kernel function operation $L(a_i \cdot a_j) = \psi(a_i)\psi(a_j)$ in low dimensional space. The following formula was used to solve the optimal hyperplane of psychological crisis early warning of college students in SVM.

$$\begin{aligned} &\max Q(\beta) = \sum_{i=1}^{m} \beta_i - \frac{1}{2}\sum_{i=1}^{m} \beta_i \beta_b b_i b_j L(a_i, a_j) \\ &s.t. \sum_{i=1}^{m} b_i \beta_i = 0, 0 \le \beta_i \le D, i = 1, 2, ..., m \end{aligned} \tag{7}$$

The corresponding decision function expression was given as follows:

$$g(a) = sign\left\{ \sum_{i=1}^{m} b_i \beta_i^* L(a_i, a) + \chi^* \right\} \tag{8}$$

For psychological crisis early warning combining PCA with SVM, different combination methods should be taken into consideration. As for the first combination method, this study extracted the principal component score b_1 of data after making PCA of raw data, which aimed to reduce data dimensions, noise and redundant information, thus improving the model's performance. The the raw data was input into the SVM model for training and prediction, which obtained prediction result b_2. Finally, the score b_1 and result b_2 was linearly combined to calculate output b, which was the final prediction result.

$$b = x_1 b_1 + x_2 b_2 \tag{9}$$

As for the second combination method, PCA of raw label data of psychological crisis was made first to extract the principal components of data, which also aimed to reduce data dimensions, noise, and redundant information. Then the principal component data was input into the SVM model. Finally, the final prediction result was obtained through SVM training and prediction.

$$b = SVM(PCA(A)) \tag{10}$$

Both methods attempted to improve the model performance by comprehensively utilizing the advantages of PCA and SVM. The first method fused the prediction results of PCA and SVM through linear combination. While the second method directly input the data, which was processed by PCA, into the SVM model, which enabled the SVM to predict more accurately based on the dimension-reduced data. The specific combination method should be chosen based on actual data, model performance, and needs.

5 Experimental results and analysis

According to the impact of dynamic embedding size on model MRR as shown in Figure 4, the model performance with different scaling parameters was analyzed using two indexes, Recall@10 and Mean Reciprocal Rank (MRR). In terms of Recall@10 index, the model achieved the best performance with 0.318 recall rate, when the scaling parameter was 6. While the recall rates with other scaling parameters (2, 4, and 8) were 0.312, 0.316, and 0.311, respectively, which were not significantly different, indicating that the model found relevant network behavior interactive prediction of college students more accurately with 6 as the scaling parameter. In terms of MRR index, the model also achieved the best performance with 0.132 MRR, when the scaling parameter was 6. While the MRRs of other scaling parameters (2, 4, and 8) were 0.13, 0.13, and 0.129, respectively, which were very close, indicating that the model had good sorting effects for the interactive prediction with 6 as the scaling parameter. Therefore, 6 was determined as the scaling parameter in order to achieve higher recall rate and sorting effects.

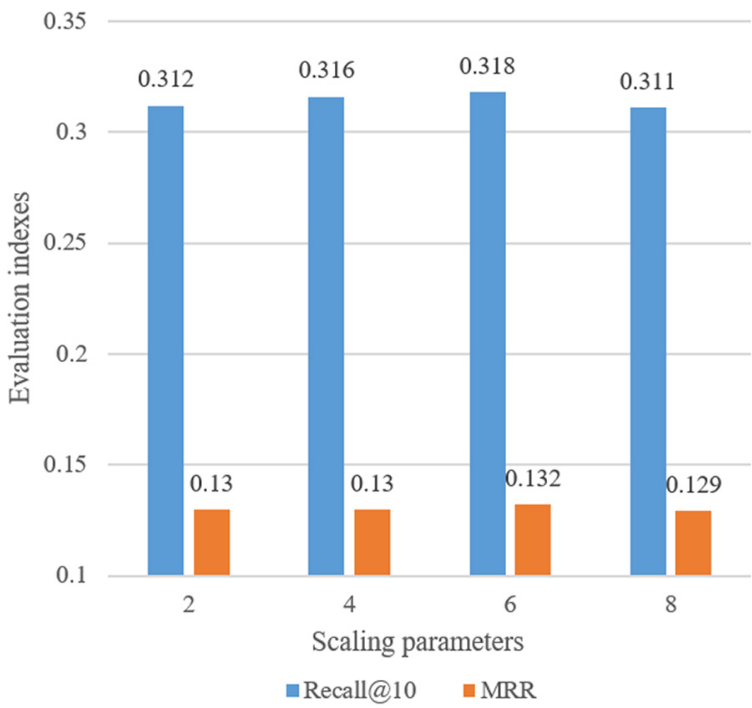

Fig. 4. Impact of dynamic embedding size on model MRR

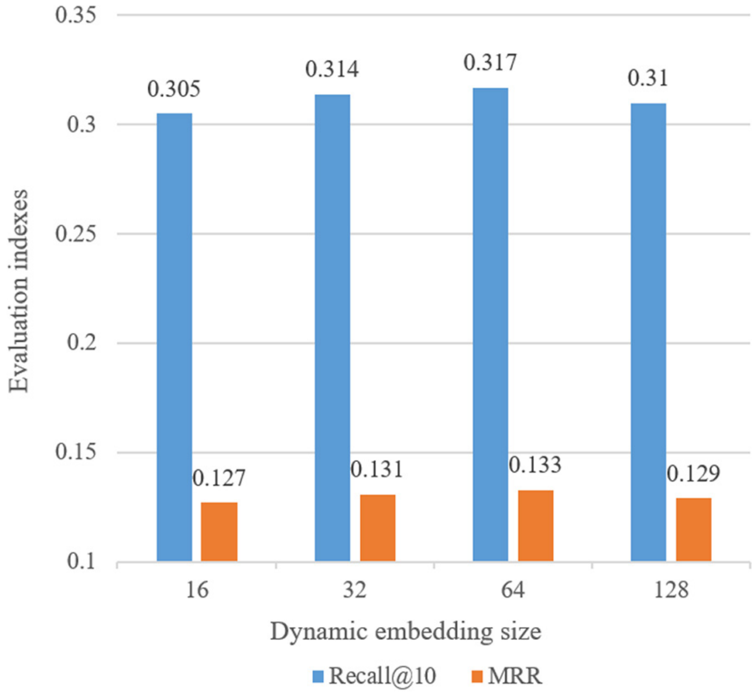

Fig. 5. Impact of dynamic embedding size on MRR

Based on the impact of dynamic embedding size on MRR as shown in Figure 5, the model performance with different dynamic embedding sizes was analyzed using two indicators, Recall@10 and MRR. In terms of Recall@10 index, the model achieved the best performance with 0.317 recall rate when the dynamic embedding size was 64, and achieved the second best performance with 0.314 recall rate when the size was 32. With 16 and 128 as the dynamic embedding sizes, the recall rates were 0.305 and 0.31, respectively, which were relatively low, indicating that the model found relevant network behavior interactive prediction more accurately with 64 as the dynamic embedding size. In terms of MRR index, the model also achieved the best performance with 0.133 MRR when the dynamic embedding size was 64, and achieved the second best performance with 0.131 MRR when the size was 32. With 16 and 128 as the dynamic embedding sizes, the MRRs were 0.127 and 0.129, respectively, which were relatively low, indicating that the model had good sorting effects for network behavior interactive prediction with 64 as the dynamic embedding size. Therefore, 64 was set as the dynamic embedding size in order to achieve higher recall rate and sorting effects.

Table 1 presents the comparative experimental results of different models. It can be seen from the table that the two indexes of MRR and Recall@10 are used to analyze the performance of different prediction models. The model in this study had the best performance, with 0.201 MRR and 0.323 Recall@10, which was close to the performance of Naive Bayes, indicating that the model in this study had good performance in dealing with network behavior interactive prediction of college students. The performance of GCN, k-Nearest Neighbor (NN), and recurrent recommender network (RRN) models was moderate, especially GCN and k-NN, with 0.234 and 0.312 comparable Recall@10. The RRN had 0.093 MRR, and 0.179 Recall@10. While the performance of Bidirectional Encoder Representations from Transformers (BERT), Time-Long Short-Term Memory (LSTM), XGBoost, DeepCoevolve, and LightGCN was relatively weak on this issue, especially XGBoost, which had relatively low 0.009 MRR and 0.012 Recall@10. In summary, the model in this study had high performance in network behavior interactive prediction of college students, while BERT, Time-LSTM, XGBoost, DeepCoevolve, and LightGCN had relatively weak performance.

Table 1. Comparative experimental results of different network behavior interactive prediction models of college students

Method	LastFM	
	MRR	Recall@10
BERT	0.061	0.121
Time-LSTM	0.070	0.142
RRN	0.093	0.179
Transformer	0.203	0.312
GCN	0.154	0.234
XGBoost	0.009	0.012
DeepCoevolve	0.019	0.043
LightGCN	0.042	0.073
k-NN	0.187	0.312
Naive Bayes	0.196	0.322
Model in this study	0.201	0.323

Table 2. Principal component score coefficient matrix

Index	F_1	F_2	F_3	F_4
Network behavior	–0.020	–0.100	0.512	–0.51
Academic performance	–0.021	–0.112	0.512	–0.049
Social relationship	0.119	–0.361	–0.129	–0.201
Living habit	0.102	–0.419	0.091	0.072
Family background	0.009	0.132	–0.089	0.741
Individual personality	0.310	–0.059	–0.039	–0.158
Psychological test	0.254	0.031	–0.034	–0.088
Historical psychological crisis event	0.192	0.151	–0.109	–0.571

Main indexes and differences of four factor score models (from F1 to F4) were analyzed, based on the provided principal component score coefficient matrix (Table 2). F1 factor focused on personal psychological and social characteristics, and the main indexes of its score model included individual personality (0.310), psychological test (0.254), historical psychological crisis event (0.192) and social relationship (0.119), with the indexes of individual personality and psychological test having the greatest impact on the factor. F2 factor focused on indexes with negative impact, such as living habit and social relationship, and the main indexes of its score model included living habit (–0.419), social relationship (–0.361), academic performance (–0.112), and network behavior (–0.100), with the living habit index having the greatest impact on the factor. F3 factor focused on the characteristics of network behavior and academic performance, and the main indexes of its score model included network behavior (0.512), academic performance (0.512), living habit (0.091), and family background (–0.089), with the indexes of network behavior and academic performance having the same greatest impact on the factor. F4 factor focused on characteristics related to family background, psychological crisis event and network behavior, and the main indexes of its score model included family background (0.741), historical psychological crisis event (–0.571), and network behavior (–0.51), with the family background index having the greatest impact on the factor.

Prediction results of the SVM model were analyzed based on the above table (Table 3). Among the five sample groups, the accuracy of the SVM model fluctuated between 90.87% (Groups 2 and 3) and 75.59% (Group 1), with an average accuracy of 84.18%. The first-type error rate fluctuated between 0.12% (Group 3) and 18.21% (Group 1), with an average error rate of 9.14%. The second-type error rate fluctuated between 0.09% (Group 2) and 9.33% (Group 5), with an average error rate of 7.33%.

Table 3. Prediction results of single model

Sample Group	1	2	3	4	5	Average Value
Accuracy	75.59%	90.87%	90.87%	81.79%	81.79%	84.18%
First-type error rate	18.21%	9.12%	0.12%	9.12%	9.12%	9.14%
Second-type error rate	9.15%	0.09%	8.98%	9.21%	9.33%	7.33%

Table 4. Prediction results of combined model

Sample Group	1	2	3	4	5	Average Value
Accuracy	90.87%	99.99%	90.87%	90.87%	81.91%	90.90%
First-type error rate	9.01%	0.13%	9.12%	0.01%	9.21%	5.50%
Second-type error rate	0.00%	0.01%	0.04%	9.13%	9.24%	3.68%

Table 4 shows the prediction results of combined model. Among the five sample groups, the accuracy of the combined prediction model fluctuated between 81.91% (Group 5) and 99.99% (Group 2), with an average accuracy of 90.90%, indicating that the model had better overall performance in network behavior interactive prediction compared with the single model. The first-type error rate fluctuated between 0.01% (Group 4) and 9.21% (Group 5), with 5.50% average error rate, indicating that this model had a lower error rate than the single model when predicting non-positive samples as positive ones. The second-type error rate fluctuated between 0.00% (Group 1) and 9.24% (Group 5), with 3.68% average error rate, indicating that this model also had a lower error rate than the single model when predicting positive samples as non-positive ones.

In summary, the constructed combined prediction model had very good overall performance in network behavior interactive prediction of college students, with high accuracy and relatively low first-type and second-type error rates. Compared with the single model, the combined prediction model improved its performance in terms of accuracy and the first-type and second-type error rates, which verified the effectiveness of combining PCA with SVM in psychological crisis early warning of college students in this study.

Table 5. Prediction results of models

Group	Logistic Regression	Traditional BP Neural Network	Single SVM Model	Model in this Study
Accuracy	70.41%	74.49%	84.18%	90.90%
First-type error rate	14.79%	10.88%	9.14%	5.50%
Second-type error rate	14.79%	14.47%	7.33%	3.68%
Total error rate	29.58%	25.35%	16.47%	9.18%

Prediction results of different psychological crisis early warning models for college students were compared and analyzed according to Table 5. It can be seen from the table that the model in this study had the highest accuracy of 90.90%, which is significantly superior to other models. The single SVM model had the second highest accuracy of 84.18%. The accuracy of traditional BP neural network and logistic regression models was 74.49% and 70.41%, respectively. The model in this study had the lowest first-type error rate of only 5.50%, which was much lower than that of other models, the lowest second-type error rate of 3.68%, which was significantly better than that of other models, and the lowest total error rate of 9.18%, which was much lower than that of other models. In summary, the model in this study had better performance in psychological crisis prediction of college students than that of other three models, in terms of accuracy, the first-type and second-type error rates, and total error rate. Therefore, this model had high predictive performance and practical value in the psychological crisis early warning of college students.

6 Conclusion

This study was about the psychological crisis prediction of college students based on big data mining of network behavior. After defining the network behavior interactive prediction of college students, the objective function of the constructed model was determined. This study presented the interactive prediction model framework and explained the working principle of the model. Then various early warning indexes were given, which needed to be comprehensively considered in the psychological crisis early warning model of college students, and the combination of PCA and SVM was applied to the construction of the early warning model, thus improving its prediction effects, generalization ability and interpretability, and reducing the overfitting risk and the difficulty of processing high-dimensional data.

Comparative experimental results of different prediction models were presented, which verified that the model in this study had high network behavior interactive prediction performance of college students. Main indexes and differences of four factor score models were analyzed, based on the provided principal component score coefficient matrix. Prediction results of single and combined model verified the effectiveness of combined PCA with SVM in psychological crisis early warning. Prediction results of different models were compared and analyzed, which verified that the model in this study had high predictive performance and practical value in the early warning. Finally, psychological crisis risk level distribution on positive and negative samples was provided, and the analysis results were presented.

7 References

[1] Huang, Y.Y. (2022). Research on psychological problems and countermeasures of contemporary college students based on data analysis. Mobile Information Systems, 2022: 3366837. https://doi.org/10.1155/2022/3366837

[2] Kassen, G., Kudaibergenova, A., Mukasheva, A., Yertargynkyzy, D., Moldassan, K. (2021). mobile and web-based support in overcoming behavioral difficulties of adolescents. International Journal of Emerging Technologies in Learning, 16(04): 69–81. https://doi.org/10.3991/ijet.v16i04.18577

[3] Lin, Y., Zhou, W. (2020). Educational strategies for coping with problems of stay up late based on the psychological characteristics of contemporary Chinese college students. In Advances in Human Factors in Training, Education, and Learning Sciences: Proceedings of the AHFE 2019 International Conference on Human Factors in Training, Education, and Learning Sciences, Washington DC, USA, 301–309. https://doi.org/10.1007/978-3-030-20135-7_30

[4] Jia, Z. (2022). Prediction of college students' psychological crisis with a neural network optimized by harmony search algorithm. International Journal of Emerging Technologies in Learning, 17(02): 59–75. https://doi.org/10.3991/ijet.v17i02.29009

[5] Wang, B., Liu, S. (2021). Prediction method of college students' psychological pressure based on deep neural network. Scientific Programming, 2021: 2943678. https://doi.org/10.1155/2021/2943678

[6] Sakka, F., Gura, A., Latysheva, V., Mamlenkova, E., Kolosova, O. (2022). Solving technological, pedagogical, and psychological problems in mobile learning. International Journal of Interactive Mobile Technologies, 16(02): 144–158. https://doi.org/10.3991/ijim.v16i02.26205

[7] Zhou, N., Ma, S., Li, X., Zhang, J., Liang, Y., Yu, C., Geng, X.M., Meng, J.B., Yuan, X.J., Cao, H.J., Fang, X. (2019). Peer contagion processes for problematic internet use among Chinese college students: A process model involving peer pressure and maladaptive cognition. Computers in Human Behavior, 90: 276–283. https://doi.org/10.1016/j.chb.2018.09.016

[8] Jia, R. (2017). An empirical analysis of college students' psychological capital interaction and pressure regulation based on SEM model. Boletin Tecnico/Technical Bulletin, 55(20): 296–302.

[9] Braasch, J., Goldman, S.R. (2006). College students' understandings of pressurized air movement: do isomorphic questions elicit isomorphic answers? ICLS 2006-International Conference of the Learning Sciences, Proceedings, 2: 892–893.

[10] Huang, M., Dong, L., Kuang, H., Jiang, Z.Z., Lee, L. H., Wang, X. (2021). Supply chain network design considering customer psychological behavior-a 4PL perspective. Computers & Industrial Engineering, 159: 107484. https://doi.org/10.1016/j.cie.2021.107484

[11] Kroesen, M., Chorus, C. (2020). A new perspective on the role of attitudes in explaining travel behavior: A psychological network model. Transportation Research Part A: Policy and Practice, 133: 82–94. https://doi.org/10.1016/j.tra.2020.01.014
[12] Yuan, L. (2020). Analysis of college students' psychological behavior and research on educational management strategies in the network information environment. In 2020 International Conference on Modern Education and Information Management (ICMEIM), 218–221. https://doi.org/10.1109/ICMEIM51375.2020.00057
[13] Ryabinin, K.V., Chuprina, S.I., Belousov, K.I., Permyakov, S.S. (2018). Visual analytics methods of the verbal behavior variability of social networks users depending on their individual psychological features. In GraphiCon 2018-28th International Conference on Computer Graphics and Vision, 167–171.
[14] Teng, C.E., He, C.Y. (2014). The research about influential factors of the new network of community based on behavior psychological of audience. Advanced Materials Research, 971: 1427–1430. https://doi.org/10.4028/www.scientific.net/AMR.971-973.1427
[15] Kim, J.W., Chock, T.M. (2017). Personality traits and psychological motivations predicting selfie posting behaviors on social networking sites. Telematics and Informatics, 34(5): 560–571. https://doi.org/10.1016/j.tele.2016.11.006
[16] Masud, S., Pinkesh, P., Das, A., Gupta, M., Nakov, P., Chakraborty, T. (2022). Half-Day Tutorial on combating online hate speech: the role of content, networks, psychology, user behavior, etc. In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 1629–1631. https://doi.org/10.1145/3488560.3501392
[17] Saleema, A., Thampi, S.M. (2020). User recognition using cognitive psychology based behavior modeling in online social networks. In Advances in Signal Processing and Intelligent Recognition Systems: 5th International Symposium, SIRS 2019, Trivandrum, India, 130–149. https://doi.org/10.1007/978-981-15-4828-4_12
[18] Yoshida, T., Hasegawa, M., Gotoh, T., Iguchi, H., Sugioka, K., Ikeda, K.I. (2007). Consumer behavior modeling based on social psychology and complex networks. In The 9th IEEE International Conference on E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007), 493–494. https://doi.org/10.1109/CEC-EEE.2007.36
[19] Jia, Z. (2022). Prediction of college students' psychological crisis with a neural network optimized by harmony search algorithm. International Journal of Emerging Technologies in Learning (iJET), 17(2): 59–75. https://doi.org/10.3991/ijet.v17i02.29009
[20] Liu, J., Shi, G., Zhou, J., Yao, Q. (2021). Prediction of college students' psychological crisis based on data mining. Mobile Information Systems, 2021: 1–7. https://doi.org/10.1155/2021/9979770

8 Author

Zhiping Jia is currently a lecturer in Hebei Chemical & Pharmaceutical College. She graduated from Hebei Normal University with a master's degree in psychology, research direction including basic psychology, education and development psychology, counseling psychology, has published a core paper in psychology, one SCI, more than 20 other provincial papers, psychological professional related utility model patents 4, 1 national invention patent. Email: jzp123004@163.com, Orcid: https://orcid.org/0000-0003-4787-4218.

Article submitted 2023-02-23. Resubmitted 2023-04-29. Final acceptance 2023-05-05. Final version published as submitted by the authors.