Hindawi

Research Article

Construction and Prediction of Students' Multiattribute Social Network Based on Psychological Big Data Analysis

Na Zhang[1] and Yuanyuan Zou [2]

[1]*Mental Health Education Center, Chengdu Agricultural College, Chengdu 611130, Sichuan, China*
[2]*College of Health Management, Shangluo University, Shangluo 726000, Shaanxi, China*

Correspondence should be addressed to Yuanyuan Zou; 209005@slxy.edu.cn

In view of the limitations of the current research on students' single attribute psychological problems. In this paper, a multiattribute social network model was constructed based on students' social data and psychological label data, and the improved MANE algorithm was used to solve the problem to predict students' psychological problems. In addition, DeepWalk and Node2vec network embedding algorithms were used to embed students' multiattribute social network, respectively, so to verify the effectiveness of the model. Finally, based on the prediction model of students' psychological problems, this paper uses Django web framework, MySQL database, and Bootstrap framework to design a personality prediction system, including data storage, algorithm training and prediction, result display, and other modules.

1. Introduction

With the rapid development of China's higher education, the number and quality of students' training are steadily improving. The state and society have also given more and more attention to higher education. However, many freshmen have been fully engaged in the college entrance examination before they enter the university, so they have little contact with the society, which makes many students unable to adapt to the great changes in their environment and lifestyle; compared with other social groups, they are more likely to have some psychological problems [1, 2]. Most of the senior students have gone through the adaptation period, but the free learning and living environment, complex interpersonal relationship, huge test pressure, and changes in the family situation still have a great impact on students' psychology.

Usually, the identification of psychological problems relies heavily on a psychological questionnaire survey. However, in the questionnaire with uneven quality, people tend to reflect better on themselves, which usually does not show a person's true inner thoughts. However, the students' quantitative strength of some problems varies from person to person, which will lead to the data being unable to describe the respondents well. Therefore, it is difficult to accurately identify students' psychological problems by using such noisy data [3]. Thanks to the rapid development of computer science, people have begun to analyze social networks by using indicators such as degree centrality and edge betweenness, which can better describe some characteristics of social networks. However, it can only express part of the information in social network, and it may contain more noise. Network representation learning (NRL) based on deep learning can solve this problem very well, which can usually capture more information in the network for more detailed analysis and identification [4]. However, most of the traditional personality analysis methods based on social network text information are based on mental lexicon and other analysis tools to make statistics of word frequency features in text information and then use machine learning method to construct a classifier to complete the training and prediction of the model. Although this method has made some progress

in effect, because it only learns the shallow statistical features in the data, the prediction effect of students' personality characteristics is limited [5, 6]; In addition, natural language processing technology based on deep learning has high requirements on the quality and quantity of data annotation, which is difficult to be applied in practice [7].

Therefore, this paper uses the multiattribute social network to better identify students' psychological problems and provides a new possible direction for the adaptive modification of network embedding technology in different data. At the same time, a prediction method for a psychological problem that integrates computer science and psychology knowledge is proposed, which can reduce the dependence of data annotation and improve the psychological characteristics of students. A multiattribute social network model was constructed based on students' social data and psychological label data, and the improved MANE algorithm was used to solve the problem to predict students' psychological problems.

2. Theoretical Basis of NRL

In network science, people usually use data structures such as an adjacency matrix to represent a network, which only stores the adjacency relationship between vertices. In fact, this simple representation cannot reflect more complex and higher-order structural relations, such as path, structure, and so on. NRL aims to represent nodes in the network in the form of vector that have the ability of reasoning in the vector space, so they can be used as the input of downstream tasks and applied to the social network.

2.1. Classification of NRL. As shown in Figure 1, the existing NRL technologies can be divided into two categories: unsupervised NRL and semisupervised NRL.

Unsupervised NRL is suitable for learning vertex representation without providing the label to which the vertex belongs. Therefore, it is considered to be a general task independent of subsequent learning, and the vertex representation is learned in an unsupervised manner. They are used as features in any vector-based algorithm for various learning tasks. According to the types of network information that can be used for learning, the unsupervised NRL algorithm can be further divided into two categories: the NRL algorithm which only retains the network structure, and the NRL algorithm which considers the vertex attributes and network structure to jointly learn and joint vertex embedding. While semisupervised NRL is suitable for representation learning when there are labels in vertices. Since vertex labels play an important role in determining the classification of each vertex and have a strong correlation with the network structure and vertex attributes, a semisupervised NRL is proposed, which uses the existing vertex labels in the network to find a more effective vector representation.

2.2. Embedding Method. In real life, network data is usually more complex, and there may be multiple links between nodes in the network. This kind of network is usually called a multiattribute network, which is composed of multiple network views. Multiattribute network analysis has become an important tool to understand the different relationships and interactions between nodes in complex systems where each view in the multiattribute network depicts the topology of a group of nodes corresponding to a specific relationship. In addition, the interaction between different attributes implies how different relationships interact on the topology of each attribute.

2.2.1. Deepwalk. Deepwalk is the first widely used graph embedding method, which uses the walk algorithm, which is a concept in graph theory. As long as nodes are connected to the graph, the whole graph can be traversed through the swimming between nodes [8, 9]. The following method is used to complete the movement of a deep walk:

$$P = P_r\left(V_i | w(v_1), w(v_2), \ldots, w(v_n)\right). \tag{1}$$

Here, P_r is the conditional probability of the expression, w represents the mapping function that embeds each node in the graph as a vector. The purpose of the formula is to estimate the selection probability of the next node according to the previous nodes in the random walk path. The potential vector representation of nodes is used as the input of the neural network, and the neural network can predict the classification of nodes according to the frequency of nodes in the path.

After obtaining a series of walk sequences composed of nodes by the random walk method, the deep walk method uses Skip-Gram algorithm to embed the graph. Just as the word2vec algorithm embeds sentences, the Skip-Gram algorithm predicts the context of each node in the graph and finally learns the vector representation of each node, and the context is in this scene, which represents the structure of the graph, the characteristics of the nodes and the adjacent nodes.

2.2.2. Node2vec. Node2vec is a famous graph embedding method similar to Deepwalk. The difference between them is that the random walk strategy in Deepwalk is completely random. Node2vec thinks that the structural equivalence and homogeneity in network structure should be preserved. The two properties correspond to width first search (BFS) and breadth first search (DFS), respectively, the former is a recursive algorithm and the latter is a deep traversal algorithm. As shown in Figure 2, the node2vec algorithm sets two superparameters p and q, which are used to control the walk strategy.

The parameter p mainly controls the BFS, while the parameter q controls the DFS.

As shown in Figure 3, BFS is suitable for learning local features, while DFS is more suitable for learning global features. For the same graph structure, using different strategies, the node2vec algorithm can return different graph embedding results. Therefore, users can adjust the parameters according to different tasks to select different strategies.
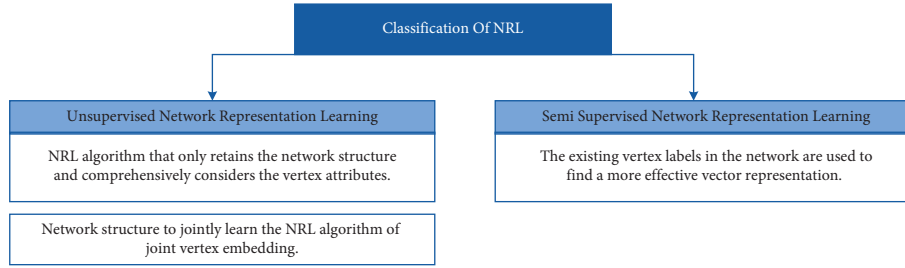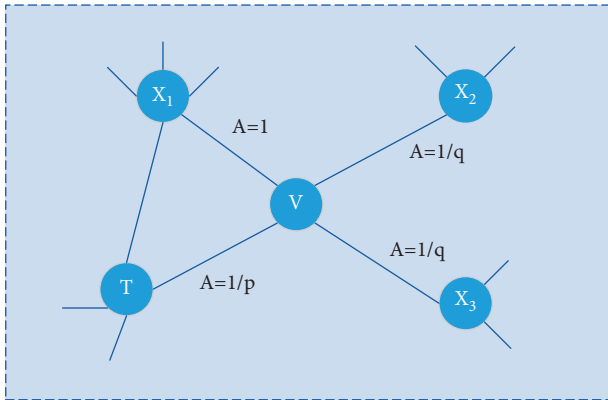
Figure 1: Classification of NRL.

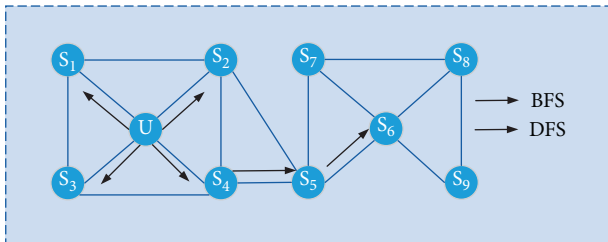Figure 2: Schematic diagram of node2vec algorithm.

Figure 3: Different walk strategies.

(1) Please list them in the order of your friends

(2) Please list in order the IDs of the students who interact with them and make you feel happy

(3) Please list the IDs of the students you think are smart in order

(4) Please list the ID of the students you will turn to when you encounter difficulties in the study

(5) Please list in order the ID of the students you will turn to when you encounter difficulties in life

(6) Please list in order the ID of the students you will actively share with when you get a piece of good news

(7) Please list in order the ID of the classmate you will share with when you learn a piece of bad news

(8) Please list in order the IDs of the students you would like to invite to join your team

Then, the answer of each question is processed and constructed into a single attribute of a social network, which has 11 attributes, reflecting the interaction between students in different social situations. By deleting the outgoing and incoming edges connected by these students, that is, taking the intersection of the nodes of the 8-attribute network, it is easy to embed and identify in the later stage, The deleted network includes 439 nodes and 221360 edges.

3. Multiattribute Social Network Model

3.1. Data Acquisition. This paper uses three types of student information, including students' social network, mental health data, and students' basic information. Before the data are collected, the purpose of data collection was informed to the students participating in the questionnaire survey, and all the data were encrypted and confused in the experiment, In order to protect the privacy of students, the student number information which may expose the student information is converted into a globally unique ID.

3.1.1. Social Network Data. In this paper, each tester is required to list 5–8 students for different social situations; that is, the degree of each node is between 5 and 8, and the listed students are orderly, that is, considering the order in which one student nominates other students, assuming that the first student nominated is the most important to the subjects, there are the following problems:

3.1.2. Psychological Label Data. In this paper, we use the possibility of psychological problems as a label for classification and use SCL-90 to assess students' possible mental health, which is a psychological test questionnaire widely used in Education. According to statistics, among 439 students (the intersection of multi-attribute social networks contains 439 nodes), 99 students may have psychological problems, the label of these students is marked as 1, and the other 340 students show that they are psychologically normal and marked as 0. Thus, the problem of identifying students' psychological problems becomes a dichotomous problem. In addition, some basic information of students (such as gender, age, and nationality) will also have a greater impact on social networks, so they are taken as additional features.

3.2. Problem Description. This paper defines the multi-attribute network as $G = (U, V, \{E^{(v)}: v \in V\})$, which consists of a node set U and an attribute set V. Each attribute

$v \in V$ contains an edge set $E^{(v)}$, $e_{i,j}^{(v)} \in E^{(v)}$ represents that node i and node $(i, j \in U)$ have an edge in attribute V.

Given a network G, the goal of multi-attribute network embedding is for each node $i \in U$ learning a low dimensional vector representation $f_i \in R^D$. In order to save the diversity of attributes as much as possible, for the $i \in U$ and each attribute $v \in V$, multi-layer network embedding will learn an intermediate representation $f_i^{(v)} \in R^{D/V}$, which only stores the information in the attribute. The final representation of a node is to splice the vectors of the node in each view, that is,

$$f_i = \oplus v \in V f_i^{(v)}, \tag{2}$$

where U is the node set and V is the attribute set;

$E^{(v)}$ is the set of edges in the attribute v

$f_1^{(v)}$ is the center embedding vector of node i in attribute v

$C_i^{(v)}$ is the context-embedded vector of node I in attribute v

f_i is the final embedding vector of node I

$\Omega^{(v)}$ denotes the pair of nodes in the attribute v

$i^{(v)}$ represents the instance of node I in attribute v

α is the hyperparameter of the first-order cooperative weight

β represents the hyperparameter of second-order cooperative weight

D represents the total dimension of each node after embedding

The resulting embedding vector not only preserves the diversity within each attribute; that is, the social network data of a student in each scene but also preserves the possible interaction between attributes. However, each attribute describes the same individual, and the inner thoughts and habits of each individual will affect his performance in different attributes. Therefore, the model tries to reconstruct the inner thoughts and habits of each student by integrating the performance in different attributes.

3.3. Prediction Model of Students' Psychological Problems. MANE algorithm is used to construct the prediction model of students' psychological problems. The idea of the MANE algorithm is mainly based on three kinds of information in multiview network: Diversity, First-order cooperation, and Second-order cooperation. Its workflow is to first run the deepwalk algorithm for each attribute to retain the unique information in each view, namely, diversity; then, the embedding vector of each view is updated according to the embedding vector of other attribute s, that is, to capture the first-order and second-order cooperation information in the multiattribute network.

The loss function of multi-attribute network embedding is shown as follows:

$$L = L_{\text{Div}} + \alpha \cdot L_{C1} + \beta \cdot L_{C2}. \tag{3}$$

Here, α and β are hyperparameters used to control the relative importance of the three relationships. When $\alpha = \beta = 0$, MANE degrades to the Skip-Gram model, such as DeepWalk, which only extracts information on a single attribute without taking advantage of any collaboration between attributes. When $\alpha > 0, \beta = 0$, MANE captures only diversity and first-order cooperation. While when $\alpha > 0, \beta > 0$, MANE can capture diversity, first-order collaboration, and second-order collaboration and can select the optimal α and β values according to the final identification results.

Literature [10] contains the specific description of L_{Div} and L_{C1}. The second-order collaboration in the MANE algorithm considers the interaction information between all views; that is, all embedded vectors in all views will be close to the context-embedded vector of the same node in other attributes. However, there may be strong or weak, promotive or antagonistic relationships among attributes. If there are antagonistic relationships and they are close to each other, some unnecessary noise may be introduced and even reduce the accuracy of recognition.

Different from the collaboration among different attributes mentioned in the MANE algorithm, the social relationship has characteristics that other networks do not have. In ordinary networks, each attribute may be relatively independent, but the multiattribute social relationship network of students is different. While the social scene in schools is relatively simple, and the new social network prefers interaction between friends. That is, most student social relationships are based on friendship or strongly related to friendship. Therefore, the loss function of second-order cooperation is optimized in this paper.

The good friend network is regarded as a special network, which has a strong correlation with all other networks that are more likely to obtain valuable information. Therefore, the first-order collaboration in MANE is modified where the good friend network is defined as v_0, and other network sets are defined as v_1, then

$$L_{C2}(\theta) = -\sum_{v \in V_1} \sum_{\left(i^{(v_0)}, j^{(v_0)}\right) \in \Omega^{(v_0)}} \log P\left(i^{(v_0)} \mid i^{(v)}; \theta\right),$$

$$= -\sum_{v \in V_1} \sum_{\left(i^{(v_0)}, j^{(v_0)}\right) \in \Omega^{(v_0)}} \log \frac{\exp\left(f_i^{(v_0)} \cdot f_j^{(v)}\right)}{\sum_{u \in U} \exp\left(f_u^{(v_0)} \cdot f_i^{v}\right)}. \tag{4}$$

4. Model Validation

4.1. Evaluation Process. Given the characteristic of label imbalance in data, and the cost of identifying abnormal students as ordinary students may be greater than that of identifying ordinary students as abnormal students, Accuracy, $F1$, and Recall are used in all experiments in this paper to evaluate the performance of the algorithm, and the model evaluation process is shown in Figure 4.
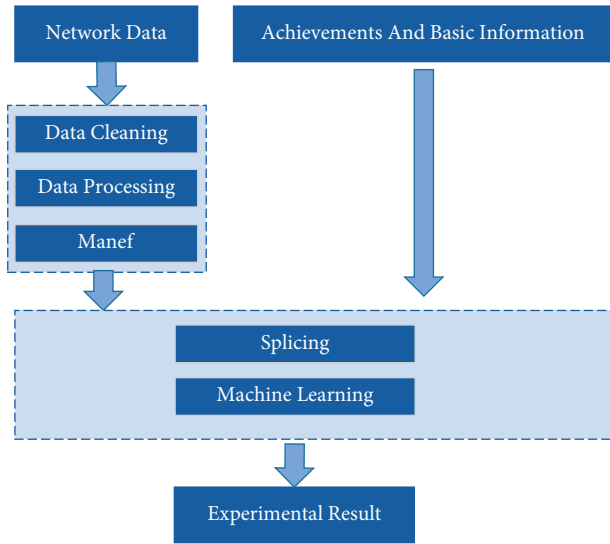
Figure 4: Process of model evaluation.

Figure 5: Comparison of different network embedding algorithms.

Suppose that each student's label (psychological abnormality) is represented by 1/0, T (true) is the correct model classification, F (false) is the model classification error, the classification result of prediction value is 1, which is represented by P (positive), and the classification result in predictive value is 0 and represented by N (negative).

4.2. Results. In this paper, a variety of DeepWalk and Node2vec network embedding algorithms are used to embed students' multi-attribute social networks, respectively. The final embedding, grades, and students' basic information generated by each algorithm are used as inputs, and DNN is used as a classifier to identify students' psychological problems. The result is shown in Figure 5.

The accuracy of network embedding vectors obtained by all network embedding algorithms has always been high, which may be due to the large imbalance of labels in the data, so this paper should focus more on the $F1$ index rather than the accuracy of the algorithm. The optimization direction of hyperparameters in this paper is to maximize $F1$, and the experimental results show that the performance of baseline algorithms such as DeepWalk and Node2vec under the $F1$ index is not as good as the algorithm proposed in this paper, which indicates that the improvement of the model in this paper is more suitable for multiview social network data used in the experiment and verifies the effectiveness of improved MANE algorithm in network embedding.

5. Design of Prediction System for Students' Personality Characteristics

5.1. Demand Analysis. The main function of the system is to help users target text information network, through carries on the personality, the network migration algorithm of knowledge map building, embedding node vector, and machine learning algorithms of embedded, finally complete personality characteristics analysis and predict the scores of the big five personalities. The functional requirements as shown in Figure 6.
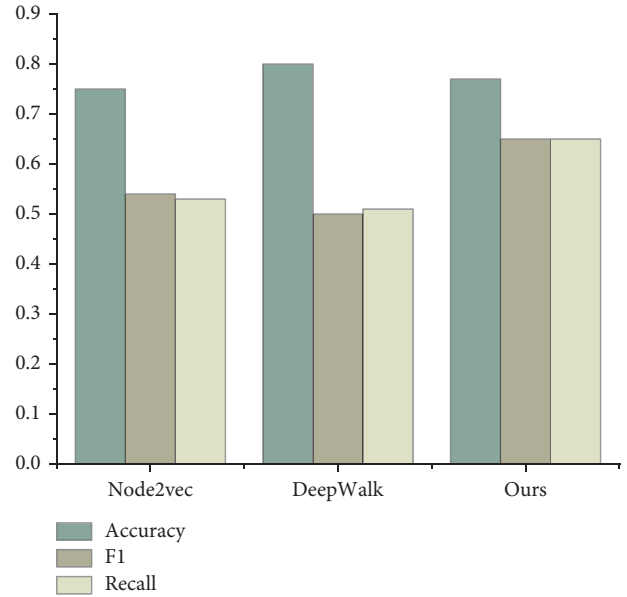
5.1.1. User Management. In order to ensure the use rights and internal data security of the system, this module is mainly responsible for the login and registration of users. For users who want to use the system to predict their personality characteristics, they need to log in first. Only after entering the correct account and password can they enter the system. For the user with wrong input information, the system will refuse to log in and give an error prompt; for users without an account, they can register the account according to the system prompts and then use the system normally. The related information of users will be stored in the background database.

5.1.2. Prediction of Personality Characteristics. As the core module of the system, this module provides users with the function of personality prediction. The detailed demand analysis of this module is as follows:

(1) *User Data Upload.* The user uploads the data set and chooses to label the training data or use the system built-in data set as the training data according to whether they have the tagging ability or not

(2) *Data Clean and Feature Extraction.* The data uploaded by users is cleaned to remove invalid data, and word frequency features are extracted to build a map

(3) *Network Walk.* The travel path is generated according to the travel strategy in the algorithm proposed above

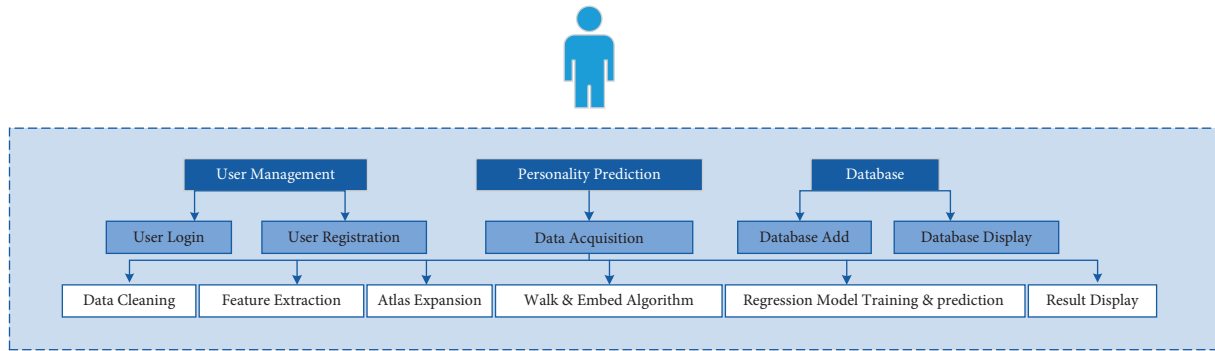(4) *Network Node Embedding.* Train the language model of the generated path to generate node vector

FIGURE 6: Functional requirements.

(5) *Regression Model Training and Prediction.* Machine learning algorithm and the generated node vector are used to train and test, and the test results are obtained

(6) *The Analysis Results Show.* After the test is completed, the test results are generated, including the big five personality score and its text description and word frequency characteristics

5.1.3. Data Management. This module mainly provides the interaction with the database for users, where they can browse the built-in data set of the system. This function mainly considers that the built-in data set of the system should be used as the training data for the users who have no annotation ability. Browsing the built-in data sets in advance can help users make better choices. In addition, the system provides the function of uploading annotation data sets. When the system administrator or other users have the annotation data of personality characteristics, they can be uploaded, which is convenient for direct use in the following prediction of personality characteristics.

5.2. Design of Systematic Architecture. According to the demand analysis of the system, this topic uses a hierarchical structure to design the system architecture and divides it into three levels: front-end interface, back-end server, and data storage layer. The three levels are responsible for different functions, and design interfaces between different levels to complete the interaction, to achieve the various functions required by the system. Through such layered design, the coupling degree between each layer of the system architecture is greatly reduced. When the system fails or other functions need to be added, only a specific layer can be modified, which greatly reduces the cost of maintenance and development and improves the scalability of the system. The overall architecture of the system is shown in Figure 7.

For each level of the system architecture, different technical means are selected according to the functional requirements. The details are as follows:.

(1) *Front-End Interface.* In the front-end interface layer, the system uses a bootstrap framework to build the front-end UI interface of the system. Bootstrap is a front-end framework developed by

Twitter, which has been widely used at present. With its excellent grid system and rich components, it can easily realize a beautiful and concise UI interface, the system uses *j*Query and Ajax to receive the user's various actions and complete the interaction function with the back-end server.

(2) *Back-End Server.* In the back-end server, the Django framework is used, which is an open source web framework written by python. Developers can easily and efficiently develop web projects through Django. It adopts MTV architecture mode; that is, model, template, and view are connected together in a loose coupling way and are responsible for different functions.

As shown in Figure 8, when a user visits the browser and makes a request, the URL controller of Django routes the request from the browser and matches it with the processing function in the view. The view function then processes the request and interacts with the part of the model responsible for the data to get the required data; next, the view function completes the processing of the request and sends the part of the result to be displayed to the template part responsible for interacting with the user. Finally, the template renders the page and returns it to the browser for a complete user request.

(3) *Data Storage Layer.* The system uses MySQL relational database to store user information and data set information, and designs the interface in the back-end server to be responsible for the interaction with the database. Among them, the data set information mainly includes the data set built in the system, which is used to provide data sources for users who do not have the ability to label, the user information mainly includes the account information of the system user, which is used to verify the login and registration. In addition, the system uses the Python library Py2neo specially responsible for docking with Neo4j in the back-end server to develop an interactive interface with Neo4j, so as to achieve the expansion, deletion, and modification and export of the knowledge graph. For exporting large-scale Atlas data, the system uses the
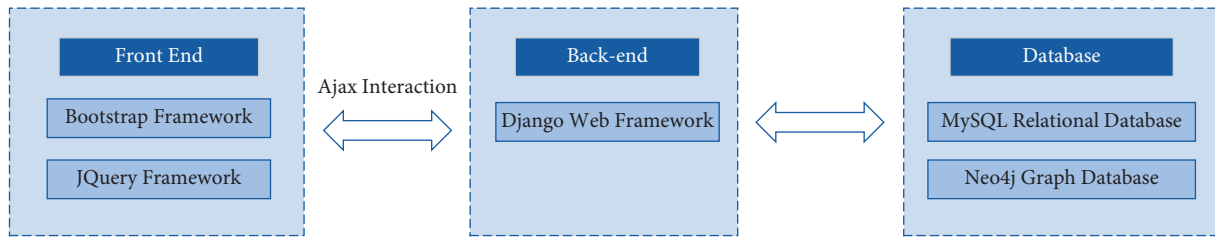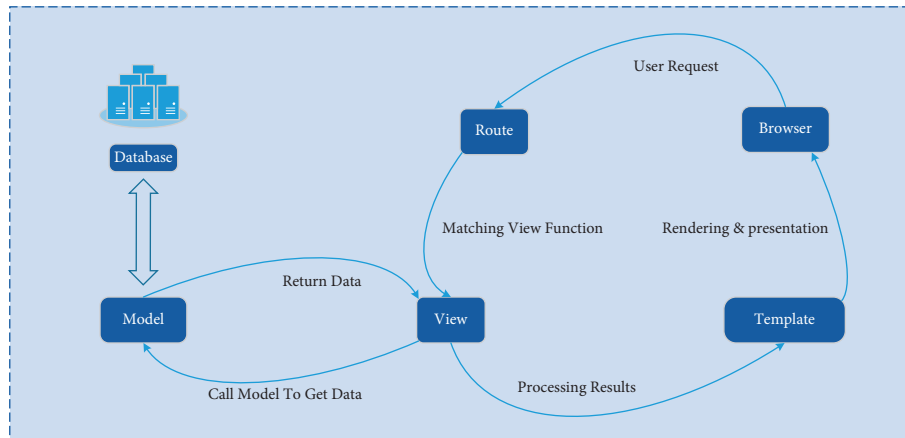
FIGURE 7: Systematic architecture.

FIGURE 8: Architecture of Django.

extension process of the Neo4j graph database and the function library APOC to implement it.

6. Conclusion

In this paper, based on the characteristics of social network and the second-order cooperation relationship between multiple attributes, a multiattribute network embedding algorithm is proposed, which solves the problem that the existing multiattribute network embedding algorithm does not consider the noise impact caused by the second-order collaboration relationship. The model can use network embedded data, student performance information, and students' basic information to identify students who may have psychological problems. Finally, the prediction system of student personality provides a visual operation platform for users, which includes user management, algorithm introduction, personality prediction, analysis results display, data display, and update to ensure that users can easily and quickly complete the task of personality prediction.

Data Availability

The data used to support the study are included in the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. Zhao, B. Qin, and T. Liu, "Sentiment analysis," *Journal of Software*, vol. 21, no. 8, pp. 1834–1848, 2010, (in Chinese).

[2] J. Tang and W. Chen, "Deep analytics and mining for big social data," *Chinese Science Bulletin*, vol. 60, no. 5-6, pp. 509–519, 2015, (in Chinese).

[3] C. Yuan, *Analysis and Recognition of Student Psychological Problems Based on Multi-View Network*, Dalian University of Technology, China, 2021, (in Chinese).

[4] J. Scott, "Social network analysis: developments, advances, and prospects," *Social Network Analysis and Mining*, vol. 1, no. 1, pp. 21–26, 2011.

[5] D. Xue, L. Wu, Z. Hong et al., "Deep learning-based personality recognition from text posts of online social networks," *Applied Intelligence*, vol. 48, no. 11, pp. 4232–4246, 2018.

[6] Y. Ding, X. Chen, Q. Fu, and S. Zhong, "A depression recognition method for college students using deep integrated support vector algorithm," *IEEE Access*, vol. 8, pp. 75616–75629, 2020.

[7] Q. Gu, C. Ju, and G. Wu, "Social Network information rec-
ommendation Model with deep autoencoder and network
representation learning," *Computer science*, vol. 47, no. 11,
pp. 101–112, 2020, (in Chinese).

[8] Q. Gu, C. Ju, and G. Wu, "Social network information rec-
ommendation model integrating deep self encoder and NRL,"
Computer science, vol. 47, no. 11, pp. 101–112, 2020, (in
Chinese).

[9] K. Berahmand, E. Nasiri, M. Rostami, and S. Forouzandeh, "A
modified DeepWalk method for link prediction in attributed
social network," *Computing*, vol. 103, no. 10, pp. 2227–2249,
2021.

[10] S. K. Ata, Y. Fang, M. Wu, J. Shi, C. K. Kwoh, and X. Li,
"Multi-view collaborative network embedding," *ACM
Transactions on Knowledge Discovery from Data*, vol. 15, no. 3,
pp. 1–18, 2021.