

## Research Article

# Analysis of Influencing Factors of Social Mental Health Based on Big Data

Yixuan Zhao<sup>1</sup> and Qinghua Tang<sup>2</sup> 

<sup>1</sup>Far East University of the Philippines, Liaocheng, Shandong 252000, China

<sup>2</sup>School of Public Health and Management, Guangxi University of Chinese Medicine, Nanning, Guangxi 530200, China

Correspondence should be addressed to Qinghua Tang; 2019163801@feu.edu.ph

Received 11 March 2021; Revised 25 March 2021; Accepted 9 April 2021; Published 19 April 2021

Academic Editor: Mian Ahmad Jan

Copyright © 2021 Yixuan Zhao and Qinghua Tang. 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.

Big data is a large-scale rapidly growing database of information. Big data has a huge data size and complexity that cannot be easily stored or processed by conventional data processing tools. Big data research methods have been widely used in many disciplines as research methods based on massively big data analysis have aroused great interest in scientific methodology. In this paper, we proposed a deep computational model to analyze the factors that affect social and mental health. The proposed model utilizes a large number of microblog manual annotation datasets. This huge amount of dataset is divided into six main factors that affect social and mental health, that is, economic market correlation, the political democracy, the management law, the cultural trend, the expansion of the information level, and the fast correlation of the rhythm of life. The proposed model compares the review data of different influencing factors to get the correlation degree between social mental health and these factors.

## 1. Introduction

Big data mainly refers to the relatively macroscopic network data generated by Internet platforms and has become an important part of big data research. It is different from microscopic big data generated by genes and brain sciences. Social science research based on big data analysis technology is of great significance for mastering political, economic, and social psychology and behavioral laws [1]. Network big data analysis technology has some advantages, that is, sample size and representation, relying on the advantages of the network platform; it can achieve measurement covering large-scale groups. This sample is close to the overall characteristics and advantages and is conducive to solving traditional research. The method represents the problem, timeliness: network big data makes it possible to track large-scale group measurements in a regular, even real-time manner. Its tracking time can be as good as yearly, monthly, daily, and hourly, or even every minute. Objectivity, network big data is based on the objective behavior data of users, such as search engine search and clicks behavior, social network likes, and forwards and

publishes content and has good objectivity of evidence. Cost-effectiveness: traditional research methods are often limited by research costs such as human resources and financial resources. It is impossible to achieve regular and real-time measurements of large-scale groups. Network big data analysis is supported by technologies such as web crawling and text analytics. It has been possible to obtain large amounts of data at a relatively low cost [2].

Research methods based on massively big data analysis have led to thinking about scientific methodology. Research does not require direct contact with the research subject, but new research findings can be obtained by direct analysis and mining of massive data, which may have spawned a new research model [3]. To this end, Turing Award winner Gray distinguishes data-intensive science from computational science and depicts the “fourth paradigm” based on data-intensive scientific research [4]. However, some researchers have doubts about big data analysis technology and believe that big data cannot replace the original research methods. It can be seen that there is still controversy about how to treat the big data analysis technology of the network.

In this paper, to overcome the problem we proposed a new framework model based on psychology which is also one of the important areas of the combination of big data and social sciences [5]. Our main contributions to this study are to present big data analysis methods that have been widely used to solve emotional psychology [6]; behavioral economics, personality psychology, and health psychology [7]; political psychology; and many other important psychological issues. So, in combination with other psychological empirical methods and for the effective features set in this study, we treat psychological research based on network big data analysis technology. The proposed model compares the review data of different influencing factors to get the correlation degree between social mental health and these factors. Unfortunately, the current systematic thinking about the methodology of online big data psychology is still very small [8].

The rest of the paper is organized as follows: In Section 2, the methods are discussed in detail. Section 3 includes big data key technologies while in Section 4, the neural network algorithm is discussed. Similarly, in Section 5, experimental results and discussion are discussed. Finally, we demonstrate a conclusion and future work.

## 2. Methods

In this section, we present the two different research methods that are used in this literature. These research techniques include cognitive neuroscience method technology and traditional research method techniques.

*2.1. Cognitive Neuroscience Technology.* Compared with modern research techniques, cognitive neuroscience techniques were formed in the 1980s and 1990s. Cognitive neuroscience techniques are formed by the combination of cognitive science and neurology. Cognitive neuroscience research attracted widespread attention from psychologists and reflections based on the perspective of psychological development [9]. Cognitive neuroscience research, mainly through modern cognitive neuroscience research techniques, such as brain wave (EEG), functional magnetic resonance imaging (fMRI), brain magnetic (MEG), transcranial magnetic stimulation (TMS), and other technologies, to reveal the law of the occurrence and development of psychology and behavior. Moreover, cognitive neuroscience has become one of the hottest cross-disciplinary fields in recent years. These fields mainly include the research on the neural mechanism of cognitive behavior, the neuroscience research of cognitive behavioral psychology theory, and the research on the theoretical model of cognitive behavioral psychological mechanism based on brain neural stimulation [10]. For example, explore the neural mechanisms of individual psychological perceptions, learning memories, attention, speech, executive control, thinking, emotions, and other psychological activities or behaviors [11].

*2.2. Traditional Research Techniques.* The traditional research techniques are divided into two main methods, that is, the questionnaire and behavioral method. The questionnaire

method is based on the respondent's self-reporting method for a series of self-reported questionnaires and uses this as data evidence to study people's psychological and behavioral patterns [12]. The main advantages of this method are targeted, quick access to data of larger populations. By designing structural, standardization, or open-ended questions directly related to the purpose of the research, the respondent's first-hand data on the problem is collected in a targeted manner. The questionnaire method can obtain data of large-scale populations more quickly through standardized operation procedures. Compared with the high research cost of cognitive neuroscience technology, the cost of the questionnaire survey is relatively low, and the sample coverage is relatively large. However, the questionnaire survey method has certain defects in subjective bias, sample size, and timeliness [13]. Network big data analysis technology has a relatively large advantage in these aspects. Poor objectivity: because the questionnaire survey method adopts a self-reporting method, there is a strong subjective bias in the research results, especially the social approval response is the most typical [14]. Similarly, the behavioral experiment method mainly involves the experimenter manually or designing different experimental conditions and observes the difference of behavioral results of the subjects under different experimental conditions, tests if the experimental conditions affect the outcome significantly [15]. The behavioral experiment method can be 83 degrees of data to explore causality. Since the experimental method controls the influence of other unrelated interference factors, it can be more reliable to judge whether the difference in behavior results is caused by the experimental conditions (i.e., independent variables), thereby causing causal inference. In view of the significance of causality in research, experimental methods also occupy an important position in scientific methodology [16]. In the behavioral experiment, the experimental person's artificial conditional intervention and the laboratory-specific environmental space may cause the experimental situation to differ from the actual environment of the subject; thus, the experiment the result is interference. Therefore, behavioral experiment methods generally face the challenge of low ecological validity. Although natural experiments can improve the authenticity of the environment, there are challenges in terms of operational feasibility and disturbance variable control. The network big data is constructed on the user's real network situation, collects objective data by not intervening in user behavior activities, so it has high ecological validity. The sample size is small and the representativeness is limited [17].

## 3. Big Data Key Technologies

*3.1. Big Data Features.* The term big data was born in 1997 and usually denotes a massive and compound collection of data. Big data has four outstanding features. These features include massive data, sparse value, multisource heterogeneity, and exponential growth. The massive data contains full sample data and ultrahigh dimensional complex data, rather than a small amount of sampled data. The sparse value of a single piece of data is extremely low and the correlation between the data is poor. The multisource heterogeneity data

sources are complex, the channels are wide, and most of them are unstructured data mixed structured data, which is difficult to sort and sort. In exponential growth, the amount of data contained in big data will change with the output index, and the traffic can reach terabytes [18].

**3.2. Big Data Core Technology.** To overcome the information threshold of big data and discover the information value of big data, its key technology is divided into three partial layers [19]. These include data platform, analysis platform, and display platform. The data platform is responsible for the collection and classification storage management of big data. The data gathered must be filtered and remarked. Marking data is always cleaned up and updated, and it contains most of the research value of massive data [9]. The analysis platform is responsible for big data calculation and analysis and is an important way to value the visualization. The transformation process requires strong computing platform support. The commonly used distributed big data computing frameworks are MapReduce, Parameter Server, and so forth. The analysis methods are commonly used for manual modeling or neural network analysis [20, 21]. Display platforms often use big data products to promote big data products, including big data research rules or value research models. The complete analysis: marking and extraction process of big data is shown in Figure 1.

#### 4. Neural Network Algorithm

The NN algorithm is an automatic calculation technique that simulates the brain study mechanism. The study covers NN organization simulation, learning algorithm, memory model, and network communication model. At present, the neural network algorithm models realized by research mainly include the feed-forward neural network, reply neural network, and time series memory neural network. Neural networks are the most efficient network model. Its outstanding feature learning capability can be broadly used in image and speech recognition. The feed-forward neural network is divided into multilayer structures, each layer consists of multiple sets of neurons, and information is input along the feed-forward layer, one-way flow transmission. Based on this feature, the feed-forward neural network can effectively extract the characteristics of data space structure. The most famous implementation methods such as perceptron and deep automatic encoder have made outstanding contributions in the field of artificial intelligence and computer vision development.

**4.1. Restricted Boltzmann Machine.** In 1986, Hinton and Sejnowski proposed the RBM is a multiplicative stochastic neural network. The network consists of visible units. Visible variables and hidden variables are binary variables, that is, 0, 1. A two-part graph is a whole network. The edges are exclusive to the exposed unit and unseen unit. As seen in Figure 2, there is no boundary relation between the visible and the hidden units:

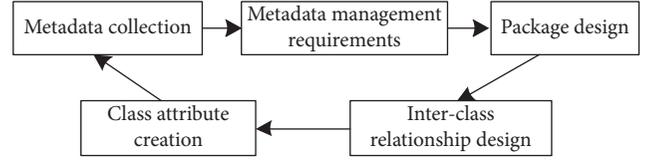


FIGURE 1: Metadata management under a three-tier architecture.

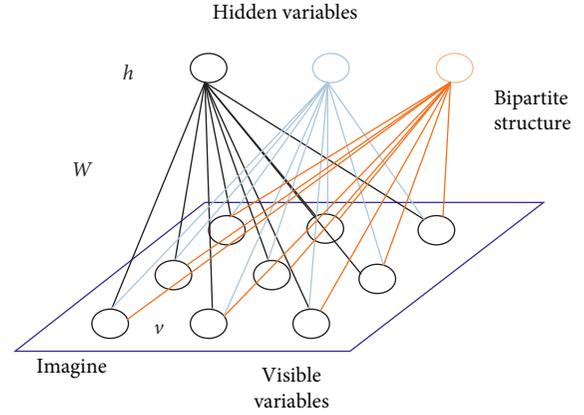


FIGURE 2: The structure of the restricted Boltzmann machine.

RBM is an energy-based model which shows visible  $v$  and hidden  $h$  energy from the joint as shown in the following equation:

$$E(v, h; \theta) = - \sum_{ij} W_{ij} v_i h_j - \sum_i b_i v_i - \sum_j a_j h_j, \quad (1)$$

where  $\theta$  is the parameter  $\{W, a, b\}$  of RBM,  $W$  is the edge between the visible and hidden units, and  $a$  and  $b$  are the bias between the visible and hidden units. Based on the joint configuration of visible variable  $v$  and hidden variable  $h$ , we can obtain the joint probability between  $h$  and  $v$ :

$$P_{\theta}(v, h) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta)), \quad (2)$$

where  $Z(\theta)$  is the normalization factor called partition function; according to function (1), function (2) can be written as

$$P_{\theta}(v, h) = \frac{1}{Z(\theta)} \exp\left(\sum_{i=1}^D \sum_{j=1}^F W_{ij} v_i h_j + \sum_{i=1}^D v_i b_i + \sum_{j=1}^F a_j h_j\right). \quad (3)$$

We hope the likelihood function of the largest observed data  $P(v)$  can be obtained by function (3):

$$P_{\theta}(v) = \frac{1}{Z(\theta)} \sum_h \exp[v^T W h + a^T h + b^T v]. \quad (4)$$

The RBM parameters we get, by maximizing  $P(v)$ , is comparable to increasing  $\log(P(v)) = L(\theta)$ :

$$L(\theta) = \frac{1}{N} \sum_{n=1}^N \log P_{\theta}(v^{(n)}). \quad (5)$$

Then, according to random gradient descent, we can get

$$(P, v | h) = \prod_i P, v_i | hP, v_i = 1 | h = \frac{1}{1 + \exp(-\sum_j W_{ij}h_j - b_i)}. \quad (6)$$

Take a sample of data and set the state of the visible variable to this sample data. Randomly initialize  $W$ . The state of the hidden variable is updated according to the first formula of equation (6); that is,  $h_j$  is set to state 1 with the probability of  $P(h_j = 1 | v)$ ; otherwise, it is set to 0. Then, for each side  $v_i h_j$ , calculate  $P_{\text{data}}(v_i h_j) = v_i * h_j$  (note that the states of  $v_i$  and  $h_j$  are both  $\{0, 1\}$ ).  $v_1$  is reconstructed according to the state of  $h$  and the second formula of the formula (6),  $h_1$  is obtained from the first formula of  $v_1$  and formula (6), and  $P_{\text{model}}(v_i h_i^1) = v_i^1 * h_i^1$  is calculated. The weight  $W_{ij}$  of the update edge  $W_{ij}$  is  $W_{ij} = W_{ij} + L * (P_{\text{data}}(v_i h_j) - P_{\text{model}}(v_i h_j^1))$ . Take the next data sample and repeat steps 1–4. The above process is iterated  $K$  times.

**4.2. Deep Belief Network.** The deep belief network (DBN) consists of a backpropagation (BP) and multilayer restricted Boltzmann machine (RBM) network. Figure 3 shows the over model. In the deep belief network, the learned output of the upper layer RBM network is used as the input of the next layer, so that each layer can better abstract the features of the upper layer and extract the data features layer by layer. The top-level BP network uses the features extracted by the RBM network as input for classification or prediction.

The RBM consists of a visible layer  $v$  and a hidden layer  $h$ , as shown in Figure 4. The visible layer is used to input feature data and the hidden layer is used for feature detectors. The nodes in the visible layer and the hidden layer have no connection with each other, that is, each node takes values independently of each other. Each node of the hidden layer can only randomly take the value 0 or 1, and the full probability distribution  $P(v, h)$  satisfies the Boltzmann distribution. The full probability distribution can be calculated by the conditional distribution  $P(v|h)$  and  $P(h|v)$ . When  $v$  is input, the hidden layer  $h$  can be obtained by  $P(h|v)$ , and after the hidden layer  $h$ , the visible layer can be obtained by  $P(v|h)$ . By adjusting the parameters, the visible layer  $v'$  obtained from the hidden layer is the same as the original visible layer  $v$ ; that is, the hidden layer is visible, another expression of the layer. Therefore, the hidden layer can be used as a feature of the visual layer input data. The joint distribution of RBM under given model parameters ( $\theta$ ) is

$$E(v, h; \theta) = \frac{\exp(-E(v, h; \theta))}{Z}, \quad (7)$$

where  $Z = \sum \sum \exp(-E, v, h; \theta)$  is the normalization factor, and the energy function is as follows:

$$(E, v, h; \theta) = - \sum_{ij} W_{ij} v_i h_j - \sum_i v_i b_i - \sum_j a_j h_j, \quad (8)$$

where  $i, j$  are nodes;  $W_{ij}$  is the connection weight between the visible layer unit and the hidden layer unit; and  $b_i$  and  $a_j$  are offsets. The BP neural network consists of three layers of

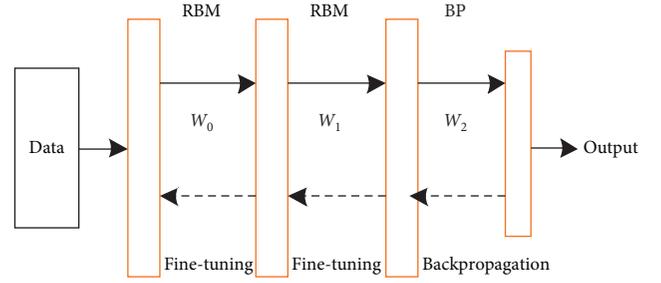


FIGURE 3: The deep belief network (DBN).

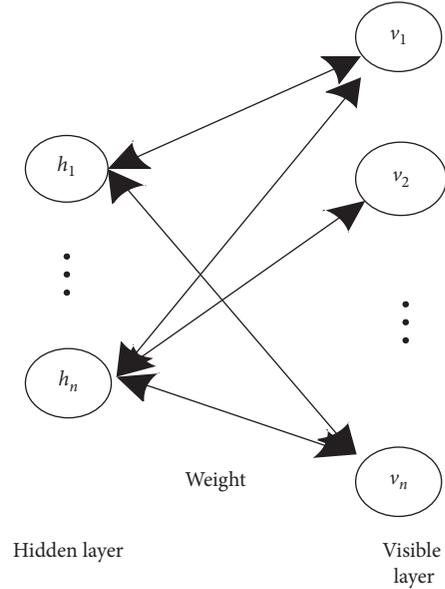


FIGURE 4: RBM structure.

neurons: input layer, hidden layer, and output layer. The structure of the BP network is shown in Figure 5. BP network in a DBN can be understood as a classifier with supervised learning.

In the BP network, the output of the hidden layer node,  $O_j = f(\sum W_{ij}x_i - a_j)$ , where  $a_j$  is the neuron threshold, and  $f$  is the excitation function, generally taking the Sigmoid function. The output of the output node  $y_k = f(\sum T_{jk}O_j - b_k)$ , where  $b_k$  is the neuron threshold and  $T_{jk}$  is the strength of the connection between the hidden layer node and the output layer node.

## 5. Experimental Results and Discussion

In this section, we analyze and discuss the results achieved by our proposed model for the microblog data of 2017 and 2018 from January to September are used as experimental data.

**5.1. Evaluation Method.** In this section, we demonstrate the measurement of the performance proposed model. The confusion metrics is one of the most widely used techniques used by several researchers for the identification of the performance result. In this paper, the correlation coefficient (Corr) and mean absolute error (MAE) are used as

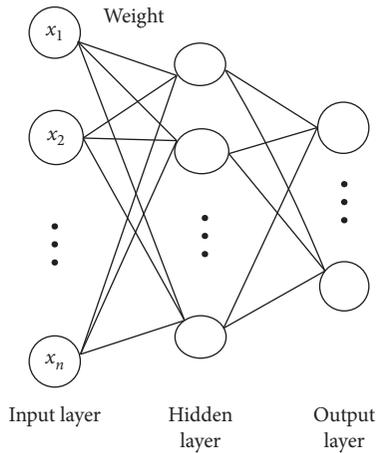


FIGURE 5: BP network structure.

evaluation indicators. The correlation coefficient is calculated as follows:

$$\text{Corr} = \frac{1}{n} \sum_{i=1}^n \left( \frac{R_i - \bar{R}}{\sigma_R} \times \frac{P_i - \bar{P}}{\sigma_P} \right), \quad (9)$$

where  $n$  is the predicted sample point;  $R_i$  and  $P_i$  are the actual mean and predicted mean of the mental health level test sample, respectively; and  $R$  is the mean and standard deviation of  $R_i$  and  $P$  is the mean and standard deviation of  $P_i$ , respectively. To calculate the average absolute error, we used the following equation:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |R_i - P_i|. \quad (10)$$

**5.2. Model Parameters.** The DBN structure is determined by the network depth. These include the number of inputs, the number of outputs, and the number of hidden layer nodes in each layer. The number of nodes in the 1<sup>st</sup> layer of RBM viewport nodes is determined by the number of input sample features. In this study, social psychology considers many influencing factors, many of the above factors; we take each of the microblog comments, that is, (1) economic market correlation degree; (2) political democracy related degree; (3) management legal system related degree; (4) cultural thoughts diverse degree; (5) degree of information expansion; and (6) rapid correlation of life rhythm. These six main influencing factors are as the characteristics of network learning; all data are manually marked and each influencing factor is also graded in a hierarchical manner so that it can be digitized as the input tensor. Each influencing factor has its own rating according to the criteria.

A major impact on model performance is the depth of the DBN network. Research shows that if the number of RBM layers increases, DBN modeling is enhanced and the hidden layer in the higher layer provides more abstract functional representation and improved network prediction performance. Overfitting may however contribute to several layers. The number of hidden nodes also influences the performance

of the model in DBN. The number of nodes is too small, the performance of the data from model mining is not good, and it can be fitted easily when too many nodes are present.

**5.3. Model Training.** The training of the DBN model is divided into two steps: pretraining and fine-tuning.

Step 1: separately train each layer of RBM networks unsupervised separately, obtain the weights of the generated models through pretraining by unsupervised greedy layer-by-layer method, and ensure that the feature information is retained as much as possible when the feature vectors are mapped to different feature spaces. The RBM training process actually determines the probability distribution that best produces the training samples by determining the weights. That is to say, a distribution is obtained such that the probability of training the sample under this distribution is the greatest.

Step 2: the BP network of the last layer of the DBN receives the output vector of the RBM as the input feature vector and supervises the classifier. Each layer of the RBM network adjusts the weights in its own layer to ensure that the feature vector mapping of the layer is optimal, and the feature vector mapping of the entire DBN is not optimal, so the BP network propagates the error information from top to bottom. Layer RBM fine-tunes a DBN network.

**5.4. Training Results.** In this paper, the microblog data of 2017 and 2018 from January to September are used as experimental data. All the data is crawled from the Internet, manually labeled, the emotional tendency is graded, and the related aspects are also graded. The grade is divided into 5 levels: 1 is the least relevant and 5 is the maximum correlation. Among them, the data from January to July of 2017 and 2018 is used as training data for the training DBN model, the data for August is used as feasibility verification data, and data of September is used as forecast test data.

In order to reasonably set the DBN's network depth, we study the impact of the DBN layer number  $\{2, 3, 4\}$  on the model prediction performance and set the number of each hidden layer node to 100. The average absolute error MAE is used as the evaluation index, and the research results are shown in Figure 6. It can be seen from Figure 6 that the DBN network depth has little effect on the accuracy of psychological prediction, and the overall three-layer structure model has the best forecast performance. In this study, DBN network depth has little effect on forecast performance, mainly due to a large amount of training data, providing sufficient data information, so that fewer RBM layers can also deeply mine data features.

Based on the above research results, we use the three-layer DBN model to further study the influence of the number of hidden layer nodes on the prediction performance of the model. The number of hidden layer nodes is set to 50, 100, and 200, respectively. MAE is also used as the evaluation index. As shown in Figure 7. It can be seen from Figure 7 that when the number of hidden layer nodes is 100,

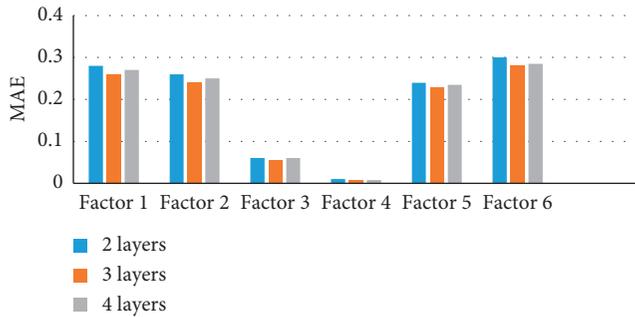


FIGURE 6: The impact of the number of DBN layers on performance.

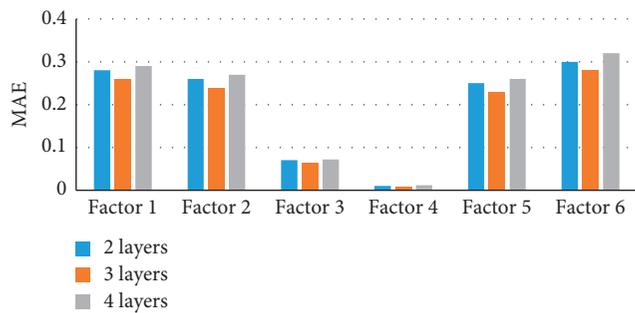


FIGURE 7: The impact of the number of hidden units of DBN on performance.

the overall prediction performance of the model is optimal. Therefore, this study will eventually adopt a three-tier DBN model with each hidden layer node set to 100.

In order to verify the accuracy of the proposed method, DBN prediction model and classical machine learning prediction model linear regression (LR), neural network (NN), support vector machine (SVM), random forest (random forest, RF), and autoregressive integral moving average model (autoregressive), the integrated moving average model (ARIMA) is compared and the results are shown in Tables 1 and 2. It can be seen from Tables 1 and 2 that the prediction performance of the DBN-based forecasting model is significantly better than other classical prediction models under the two evaluation indexes of correlation coefficient and average absolute error. This indicates that compared with other classical forecasting methods, the deep learning-based forecasting model can deeply mine the input sample features, extract the main factors affecting the mental health level, and reduce the influence of noise in the sample, thus having higher forecast accuracy.

Considering that Weibo has different differences in the hot topics in different time periods, in order to further verify the performance of the deep learning prediction model based on different environments, this paper uses the 2017 data as training data for 2018. The February and July comments were tested for predictive results as shown in Tables 3–6.

TABLE 1: Correlation coefficient of prediction results of September data.

| Model | F1    | F2    | F3    | F4    | F5    | F6    |
|-------|-------|-------|-------|-------|-------|-------|
| DBN   | 0.764 | 0.726 | 0.727 | 0.749 | 0.743 | 0.772 |
| ARIMA | 0.322 | 0.041 | 0.004 | 0.105 | 0.289 | 0.196 |
| NN    | 0.628 | 0.588 | 0.688 | 0.715 | 0.702 | 0.724 |
| RF    | 0.555 | 0.584 | 0.6   | 0.609 | 0.552 | 0.796 |
| LR    | 0.622 | 0.684 | 0.682 | 0.649 | 0.604 | 0.691 |
| SVM   | 0.524 | 0.67  | 0.647 | 0.579 | 0.688 | 0.687 |

TABLE 2: Average absolute error of prediction results of September data.

| Model | F1      | F2     | F3     | F4     | F5     | F6    |
|-------|---------|--------|--------|--------|--------|-------|
| NN    | 39.146  | 28.539 | 6.079  | 29.494 | 24.571 | 0.517 |
| RF    | 44.024  | 31.618 | 6.707  | 29.603 | 23.667 | 0.474 |
| DBN   | 24.333  | 27.805 | 5.811  | 26.974 | 22.58  | 0.403 |
| LR    | 42.277  | 31.202 | 6.686  | 30.581 | 23.615 | 0.424 |
| SVM   | 43.818  | 32.799 | 6.11   | 32.286 | 27.03  | 0.437 |
| ARIMA | 125.167 | 50.023 | 25.776 | 90.503 | 41.952 | 1.128 |

TABLE 3: Correlation coefficient of prediction results of January and February data.

| Model | F1    | F2    | F3    | F4    | F5    | F6    |
|-------|-------|-------|-------|-------|-------|-------|
| LR    | 0.652 | 0.649 | 0.543 | 0.631 | 0.588 | 0.63  |
| SVM   | 0.623 | 0.639 | 0.623 | 0.585 | 0.522 | 0.538 |
| ARIMA | 0.039 | 0.141 | 0.222 | 0.011 | 0.101 | 0.386 |
| DBN   | 0.696 | 0.744 | 0.718 | 0.698 | 0.712 | 0.773 |
| NN    | 0.536 | 0.683 | 0.648 | 0.622 | 0.652 | 0.631 |
| RF    | 0.522 | 0.654 | 0.503 | 0.522 | 0.522 | 0.575 |

TABLE 4: Average absolute error of prediction results of January and February data.

| Model | F1     | F2    | F3     | F4     | F5      | F6     |
|-------|--------|-------|--------|--------|---------|--------|
| RF    | 33.248 | 0.715 | 8.246  | 47.215 | 52.153  | 29.234 |
| LR    | 28.268 | 0.561 | 7.363  | 35.326 | 45.189  | 26.404 |
| ARIMA | 59.234 | 1.245 | 31.244 | 99.135 | 132.241 | 38.281 |
| DBN   | 24.234 | 0.491 | 6.139  | 28.351 | 27.333  | 24.891 |
| SVM   | 32.971 | 0.743 | 8.917  | 43.726 | 50.134  | 30.862 |
| NN    | 29.184 | 0.693 | 7.623  | 41.257 | 46.251  | 27.238 |

TABLE 5: Correlation coefficient of prediction results of July data.

| Model | F1    | F2    | F3    | F4    | F5    | F6    |
|-------|-------|-------|-------|-------|-------|-------|
| LR    | 0.614 | 0.674 | 0.638 | 0.603 | 0.671 | 0.683 |
| RF    | 0.543 | 0.597 | 0.596 | 0.546 | 0.572 | 0.701 |
| ARIMA | 0.314 | 0.005 | 0.106 | 0.275 | 0.031 | 0.181 |
| DBN   | 0.752 | 0.719 | 0.737 | 0.732 | 0.713 | 0.766 |
| SVM   | 0.513 | 0.634 | 0.568 | 0.679 | 0.665 | 0.679 |
| NN    | 0.613 | 0.679 | 0.707 | 0.695 | 0.589 | 0.715 |

TABLE 6: Average absolute error of prediction results of July data.

| Model | F1    | F2      | F3     | F4     | F5     | F6     |
|-------|-------|---------|--------|--------|--------|--------|
| RF    | 0.563 | 45.134  | 7.105  | 41.236 | 26.682 | 31.019 |
| LR    | 0.492 | 42.418  | 6.782  | 37.072 | 25.162 | 31.28  |
| ARIMA | 1.108 | 124.249 | 26.023 | 52.074 | 43.128 | 92.105 |
| DBN   | 0.428 | 26.291  | 5.924  | 28.285 | 23.297 | 28.284 |
| SVM   | 0.537 | 43.904  | 7.097  | 39.287 | 27.137 | 33.281 |
| NN    | 0.526 | 40.017  | 6.972  | 38.186 | 25.459 | 30.231 |

## 6. Conclusion

In this paper, we proposed a new framework that discusses the application of big data processing technology in social psychology. The proposed method is based on the deep belief network which establishes six major influencing factors that include the degree of economic market relevance of microblog commentary data, political democracy correlation, management legal system, cultural ideological diversification, information expansion, the rapid correlation of life rhythm, and the statistical model. Using the big data of the comments to train the model, fully explore the semantic features in the big data, and realize the mining of emotions and social psychology based on the commentary big data. By comparing with the classical machine learning judgment method based on the correlation and the average absolute error evaluation index, the validity of the DBN model in the mining of social psychological impact is verified. Research shows that a deep learning prediction approach can better overcome the weaknesses of conventional approaches, particularly in the case of big data, which can further explore the importance of big data in the heart and increase the implementation impact of big data comments.

Future big data research needs to deal with the relationship between data and theory. On the one hand, data-driven evidence can validate or correct existing theories, and as data-driven evidence continues to accumulate, it is expected to further refine innovative theories. On the other hand, the new theory can further provide guidance for subsequent empirical research. The combination of data driving and theoretical driving is conducive to the benign development of the mutual promotion of theory and data.

## Data Availability

Data sharing is not applicable to this article as no datasets are generated or analyzed during the current study.

## Conflicts of Interest

All authors declare that they have no conflicts of interest.

## References

- [1] J. A. Foley, N. Ramankutty, K. A. Brauman et al., "Solutions for a cultivated planet," *Nature*, vol. 478, no. 7369, pp. 337–342, 2011.
- [2] M. Herrero, P. K. Thornton, A. M. Notenbaert et al., "Smart investments in sustainable food production: revisiting mixed crop-livestock systems," *Science*, vol. 327, no. 5967, pp. 822–825, 2010.
- [3] P. J. Ericksen, J. S. I. Ingram, and D. M. Liverman, "Food security and global environmental change: emerging challenges," *Environmental Science & Policy*, vol. 12, no. 4, pp. 373–377, 2009.
- [4] K. E. Giller, "Can we define the term 'farming systems'? a question of scale," *Outlook on Agriculture*, vol. 42, no. 3, pp. 149–153, 2013.
- [5] C. Klapwijk, M. Van Wijk, T. Rosenstock, P. Van Asten, P. Thornton, and K. Giller, "Analysis of trade-offs in agricultural systems: current status and way forward," *Current Opinion in Environmental Sustainability*, vol. 6, pp. 110–115, 2014.
- [6] D. Harris and A. Orr, "Is rainfed agriculture really a pathway from poverty?" *Agricultural Systems*, vol. 123, pp. 84–96, 2014.
- [7] H. Todd, "Women climbing out of poverty through credit; or what do cows have to do with it?" *Livestock Research for Rural Development*, vol. 10, no. 3, pp. 1–9, 1998.
- [8] H. M. J. Udo, H. A. Aklilu, L. T. Phong et al., "Impact of intensification of different types of livestock production in smallholder crop-livestock systems," *Livestock Science*, vol. 139, no. 1-2, pp. 22–29, 2011.
- [9] X. Yu, Y. Chu, F. Jiang, Y. Guo, and D. Gong, "SVMs classification based two-side cross domain collaborative filtering by inferring intrinsic user and item features," *Knowledge-Based Systems*, vol. 141, pp. 80–91, 2018.
- [10] F. Khan, A. Ur Rehman, J. Zheng, M. A. Jan, and M. Alam, "Mobile crowd sensing: a survey on privacy-preservation, task management, assignment models, and incentives mechanisms," *Future Generation Computer Systems*, vol. 100, pp. 456–472, 2019.
- [11] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proceedings of the MobiCom '12: The 18th Annual International Conference on Mobile Computing and Networking*, pp. 173–184, Istanbul, Turkey, August 2012.
- [12] X. Hu, "Social drive: a crowdsourcing-based vehicular social networking system for green transportation," in *Proceedings of the ACM MSWiM '13: 16th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 85–92, Barcelona, Spain, November 2013.
- [13] M. Usman, M. A. Jan, X. He, and J. Chen, "A survey on big multimedia data processing and management in smart cities," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–29, 2019.
- [14] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: a crowdsensing-oriented mobile cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, 2013.
- [15] D. Huang, T. Xing, and H. Wu, "Mobile cloud computing service models: a user-centric approach," *IEEE Network*, vol. 27, no. 5, pp. 6–11, 2013.
- [16] X. Ge, T. Han, Y. Zhang et al., "Spectrum and energy efficiency evaluation of two-tier femtocell networks with partially open channels," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1306–1319, 2014.
- [17] S. Y. Sohn and A. S. Lee, "Bayesian network analysis for the dynamic prediction of early stage entrepreneurial activity index," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4003–4009, 2013.
- [18] X. Yu, J. Yang, and Z. Xie, "Training SVMs on a bound vectors set based on Fisher projection," *Frontiers of Computer Science*, vol. 8, no. 5, pp. 793–806, 2014.

- [19] X. Yu, F. Jiang, J. Du, and D. Gong, "A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains," *Pattern Recognition*, vol. 94, pp. 96–109, 2019.
- [20] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: a recommender system for finding passengers and vacant taxis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 10, pp. 2390–2403, 2013.
- [21] M. Babar, F. Arif, M. A. Jan, Z. Tan, and F. Khan, "Urban data management system: towards Big Data analytics for Internet of Things based smart urban environment using customized Hadoop," *Future Generation Computer Systems*, vol. 96, pp. 398–409, 2019.