News Dissemination Information Model and User Privacy Protection Method Based on BP Neural Network

Jingjing Guo*, Jianqiang Wang
School of Culture and Education, Hennan Institute Of Economics And Trade, Zhengzhou, Henan 450000, China
E-mail: guojingjing202306@163.com
*Corresponding author

Keywords: BP neural network, news spread, privacy protection, transformer

Received: April 9, 2024

Online social networks are widely used as the main way of news dissemination, but the dynamic information dissemination process in online social networks often requires more work to predict and control user privacy accurately. A novel dissemination information model and user privacy protection method based on BP neural network is proposed. First, in constructing a neural network, it is necessary to calculate the network weight vector for the training sample set. Secondly, to ensure that the private information of the neural network learning model is not leaked, this paper proposes to allocate the weight vector to all participants so that each participant has part of the private value of the weight vector. In addition, a secure multi-party computing protocol is used to ensure the safety of the intermediate and final weights of the neural network. Ensure the rationality of information dissemination and the security of user privacy. Experimental results show that the proposed algorithm has more advantages in execution time and accuracy error than traditional non-privacy protection algorithms.

Povzetek: Predstavljen je model za zaščito zasebnosti uporabnikov in razširjanje novic, ki temelji na BP nevronski mreži. Model uporablja varni večstranski računski protokol za zaščito vmesnih in končnih tež nevronske mreže, s čimer izboljšuje natančnost in učinkovitost razširjanja novic ter varuje zasebnost uporabnikov v porazdeljenih okoljih.

1 Introduction

Recently, the rapid development of Web2.0 provides a powerful impetus for the development of network media and new challenges and models for the public opinion dissemination of web community themes [1]. The emergence of a large number of social networks has brought great convenience to people's daily information acquisition [2]. The continuous interaction of social network users makes all kinds of information spread rapidly on the social network, and its extremely large social graph easily magnifies the spread scope and influence of data [3]. These characteristics make social networks an important platform for expressing public opinions and releasing information.

Building information transmission models based on real networks can be used to reflect the process of information transmission in social networks and predict the trend of information transmission in the future, which will help researchers better understand the rules of information transmission and provide theoretical support for other research based on information transmission [4], [5], [6].

Most existing recommendation systems spread and recommend for users' personalized needs according to the relationship between social networks and provide personalized recommendations for users' interest information and past behavior characteristics. The recommendation strategies mainly include mining based on association rules, semantic analysis based on content, collaborative filtering, combined recommendation algorithm, etc. Social recommendations have emerged with the rise of blogger.com, Facebook, and Twitter.

Besides, with the development of the Internet and mobile terminals, people are entering the real mobile information age. As an important mobile terminal, the phone is closely related to people's life and study and gradually becomes irreplaceable. As the most popular and widely used social service software in China, WeChat allows users to spread product information (including text or pictures, etc.) to their circle of friends through WeChat and achieve effective information dissemination and profit. Due to mobility and flexibility, WeChat information chain communication is more flexible than traditional Internet sales. WeChat information chain communication is characterized by simple operation, less investment, stable customers, and rapid spread.

Literature [7] systematically analyzes the socialization recommendation model of news information, in which a one-dimensional vector is used to depict the user's interest. The ith element in the vector 0 is that the user is not interested in category i, but 1 represents the user's interest. The success rate of recommendation is related to the structure of the interpersonal network. The higher the similarity of interest, the higher the recommendation probability. To improve the success rate of recommendation, better neighbors (with higher

similarity) are recommended to users as news sources by analyzing users' reading preferences and news dissemination patterns. For the sake of discussion, the disseminator of the news is called the leader and the receiver is called the follower. The leader-follower relationship among users continues to evolve, and the final system tends to be in a steady state with high similarity and a high recommendation success rate.

For disseminating public opinion in Web communities, typical domestic and foreign-related research work mainly includes the sznzjd model [8]. This model specifically studies the law of public opinion communication from a complex network system perspective—A small-world model. For example, literature [9] has used the small-world model to construct interpersonal network topology and introduced individual psychological factors and external media influence to establish the public opinion and communication model—propagation model combined with dynamics. For example, literature [10] divides the evolution of public opinion into two stages and studies the development of public opinion transmission using the idea of the dynamics of infectious diseases. Our survey shows no mature research results on theme-oriented public opinion communication based on Web2.0.

However, with the rapid development and popularity of the Internet, network media has become the mainstream channel of information dissemination. People can browse information on the Internet conveniently and quickly with mobile phones anytime and anywhere to obtain the information they are interested in [11]. Therefore, almost all traditional media, such as People's Daily, CCTV News, the Paper, and so on, have transferred information dissemination to the Internet in a new mode, and many new ways of information dissemination spawned by the Internet have also emerged in large numbers. WeChat public accounts, Kan-Yi-Kan, NetEase News, Tencent News, and so on are all carriers of information dissemination that users read greatly. Because of this, studying the information transmission mechanism of online media can help people better manage and control the information transmission on the network, so it is very necessary for the governance of online public opinion.

The news dissemination information model based on BP neural network proposed in this article, combined with user privacy protection methods, not only protects the privacy of participants in the scenario of horizontal partitioning of data, but also improves the efficiency of model learning in a distributed environment. The horizontal partitioning of data allows each participant to have a subset of the dataset, thus avoiding the need for data sharing. By combining privacy data measurement and rating models, it ensures that no participant's private data is leaked during the model training process. This is in stark contrast to traditional centralized learning algorithms, which typically require all datasets to be centralized in one place, increasing the risk of data leakage.

Although the data is distributed, the proposed method allows all participants to jointly construct a neural network learning model. Through a secure weight update allocation mechanism, each participant can update their weight locally without exposing their original data. This not only protects privacy, but also reduces the cost of data transmission, thereby improving learning efficiency in distributed environments. By introducing news quality factors, the model proposed in this article can more accurately describe the depth and breadth of news dissemination. This in-depth understanding of the dynamics of news dissemination, combined with privacy protection mechanisms, makes the model more applicable in real-world news dissemination scenarios.

2 Related works

2.1 Social network information dissemination

Many scholars have done research from different perspectives to explore the mechanism and influencing factors of online media information transmission, which provides good theoretical support to help people understand the evolutionary characteristics of online information transmission [12], [13], [14]. Since the nonlinear dynamics model can effectively describe the evolution of information transmission mode over time, many scholars have used it to study the transmission characteristics of online media information. In the field of information transmission, the infectious disease model is very mature. And the basic models include SI (Susceptible-Infected) and SIR(Susceptible-Infected-Recovered) [15], where S represents the susceptible node without immunity, I represent the node that has been infected by the virus and can infect other nodes, and R represents the node that is immune to the virus. In practical applications, nodes in different S, I, and R states are assigned different meanings due to different application scenarios. For example, literature [16] uses the SIR model to study the law of microblog information diffusion. It defines S, I, and R as users who have not forwarded information, users who have forwarded information, and users who are submerged. In literature [17], S, I, and R are defined as communicators, ignorant, and rationalists in the rumour propagation model. In addition, many evolutionary models, such as SIRS, SIRA, and SCIR, have been generated based on the basic model, depending on the details of the study. These models do not consider the latent state when the node changes from S to I state. Therefore, the latent state is introduced into the SIR model, generating SEIR series models.

These existing models respectively explain the mechanism of network information dissemination from different perspectives. Literature [18] reveals the basic rules of the cycle, climax, and submerging of microblog information dissemination through empirical analysis and simulation by collecting Sina Weibo data, and the research results are of guiding significance to the management practice of social network information dissemination. Considering the bandwagon effect in social networks, literature [19] has studied the influence of the bandwagon effect on rumor propagation in social networks by improving the traditional SIR model. The results show that

the number of people forwarding rumors is directly proportional to the scale of the bandwagon effect, and expanding the scale of social networks is beneficial to weaken the influence of the bandwagon effect on rumor transmission. Literature [20] constructed an improved SIR rumor propagation model with a conformity effect by considering the characteristics of rumors themselves and the number of individuals who believe and spread rumors and analyzed the dynamic state of the model. The experimental results show that the bandwagon effect accelerates rumor propagation, and this phenomenon is more obvious in the scale-free network. Based on the SIR model in infectious disease dynamics, the user status node was added to the literature [21], and the blockchain social network information transmission model $was constructed to truly reflect the rules of social network information transmission in the blockchain environment.

Table 1: Summary table

Serial Number	Citation	Research method	Propagation model	Key findings/contributions	Limitations
1	[12]	theoretical support	various	Provided a theoretical framework for network information dissemination	Lack of empirical research
2	[13]	theoretical analysis	SIR	Explored the application of SIR model in network information dissemination	No privacy issues involved
3	[14]	empirical research	SEIR	Revealed the evolutionary characteristics of network information dissemination	Privacy protection policy not involved
4	[15]	Infectious disease model	SI, SIR	Introduced basic infectious disease models	Lack of adaptability to modern social networks
5	[16]	SIR model	Weibo dissemination	Analyzed the laws of Weibo information dissemination	Not considering privacy breach risk
6	[17]	SIR model	Rumor spreading	Defined the node states in rumor propagation	Privacy protection methods not discussed
7	[18]	Empirical analysis and simulation	Sina Weibo data	Revealed the cycle and pattern of Weibo information dissemination	Lack of privacy protection measures
8	[19]	Improvement of SIR model	Rumor spreading	Studied the impact of following the trend on the spread of rumors	Privacy protection policy not involved
9	[20]	Integration effect model	Rumor spreading	Analyzed the integration effect in rumor dissemination	Privacy protection methods not discussed
10	[21]	SIR model extension	Blockchain social network information	Built an information dissemination model in the blockchain environment	Lack of quantitative assessment of privacy breaches

2.2 Privacy protection

Recently, the information industry, such as mobile Internet and big data computing platforms, has brought great convenience to people's lives, and many service-oriented Internet industries have emerged at the historic moment. While these industries provide services to users, huge amounts of data and information flow. Take the online car-hailing platform as an example; users' personal information, travel information, driver information, vehicle information, and other data constantly interact among users, platforms, and drivers. Data is often presented for service satisfaction in the interaction process, while data security is often ignored [22]. Huge amounts of data and information flow between usually contain a huge information value, not the lack of data and information related to privacy involved, even if it's just a simple comment on the network platform about the car may cause personal privacy leak, how to guarantee the safety of this type of data privacy information is a real problem need to solve.

Over the years, many researchers have done many works on how to protect private data. Based on data disturbance, data anonymity, and other strategies, some effective privacy protection models and methods are proposed, for example, the K-anonymity model, the L-diversity model, and differential privacy protection technology. The proposal and development of these privacy protection technologies have laid a solid foundation for protecting private data. However, in practical applications, they are still restricted by many types of privacy data and complex privacy application scenarios; among them, the difficulty of identifying private data is particularly prominent. As privacy is a very abstract concept, the category of privacy varies greatly in different privacy scenarios and subjects, and it is difficult to form a set of general privacy definition standards, which causes great difficulties in the identification of private data. At present, the carrier of personal information is often the massive data flow between networks; if the data that needs privacy protection cannot be successfully selected in the big data environment, the application of privacy protection technology indiscriminately in the whole network data flow will cause huge costs in time and space. Scientific and efficient privacy measurement and classification of data is a necessary prerequisite to solving the difficult problem of privacy recognition.

In the absence of universal privacy standards, it is difficult to make a "yes" or "no" decision on the privacy of a piece of data; one way to do this is to replace privacy standards with measures and grading rules; the evaluation theory is applied to privacy measurement and classification, that is, the privacy measurement and classification of a single piece of data are transformed into the privacy evaluation of data sets, by selecting privacy elements to be measured as evaluation indicators, privacy measurement and classification of single data in the data set are carried out based on relevant evaluation methods and the overall privacy status of the data set as the standard. In this way, the "barrier" of unclear privacy definition standards is bypassed. Meanwhile, privacy measurement and classification of a single record based on data sets can better reflect the privacy importance of a particular record in the immediate situation and provide a basis for implementing privacy protection technology and strategy. But there are still two key issues:1) The efficiency problem is caused by the variety and complexity of privacy measurement elements due to the broad concept of privacy.2) Due to the diversity of privacy application scenarios and the uncertainty of subjective factors of private owners, it is difficult to determine the weight of privacy measurement elements.

3 Model design

3.1 Back propagation neural network (BP)

BP has a transmission structure is a multi-layer feedforward network trained according to the error backpropagation. It contains the input layer, hidden layer, and output layer, and it includes two learning phases, forward propagation and backpropagation, which are repeated until the requirements are met. In the forward phase, the input signal passes through the input layer to act on the hidden layer and is output from the output layer. In the back phase, the connection weight values of nodes at each layer are corrected using the errors between the expected result and the actual output, and then forward propagation is carried out. The structure of the three-layer BP network is shown in Figure 1. BP network has a strong nonlinear mapping function, can be self-trained, adjusted, and improved according to the input and output, has the characteristics of self-learning and self-adaptation, and can predict the outcome more accurately, objectively, and effectively for news information dissemination, evaluation, and early warning [23].

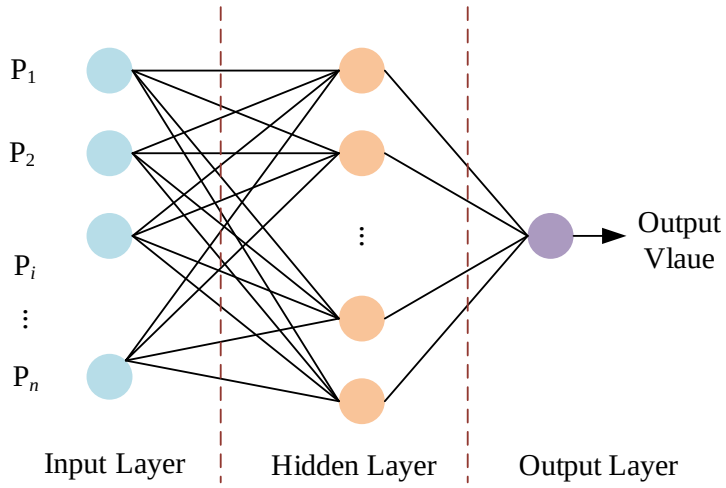

Figure 1: BP network structure diagram

The biggest advantage of BP networks is their ability to learn and store many input-output relations without revealing such mathematical links in advance, including 2 processes, forward of signals and backward phase of errors. In the forward step, the input acts on the output node through the implied layer and undergoes a nonlinear transformation to produce the output signal. If the actual output does not match the desired result, it is transferred to the backward propagation process of the error. Error back-propagation is to back-propagate the output error through the implied layer to the input layer by layer and apportion the error to all units in each layer, using the error signal obtained from each layer as the basis for adjusting the weights of each unit.

The BP neural network algorithm is executed in the following steps.

(1) Set initial values for the weight coefficients W_{ij}, where $W_{i,n+1} = -\theta$.

(2) Input a sample and the corresponding desired output $Y = (Y_1, Y_2, \ldots Y_n)$.

(3) For the output of the *i*-th neuron in the kth layer, the calculation is shown in Equation (1).

$$\begin{cases} U_i^k = \sum_{j=1}^{n+1} W_{ij} X_j^{k-1} \\ \sum X_{n+1}^{k-1} = 1 \\ W_{i,n+1} = -\theta \\ X_i^k = f(U_i^k) \end{cases} \quad (1)$$

(4) For the output layer with k = m, the learning error of each layer is calculated as shown in Equation (2). For the rest of the layers, the calculation is shown in Equation (3).

$$d_i^m = X_i^m (1 - X_i^m)(X_i^m - Y_i) \quad (2)$$

$$d_i^k = X_i^k (1 - X_i^k) \sum_i W_{li} d_i^{k+1} \quad (3)$$

(5) Correct the weight coefficient and the threshold θ. Use the weight correction as shown in equation (4).

$$W_{ij}(t+1) = W_{ij}(t) - \eta d_i^k . X_j^{k-1} + \alpha \Delta W_{ij}(t) \quad (4)$$

(6) When the individual weight coefficients of each layer are derived, the requirements can be discriminated according to the given index, and the convergence accuracy of the expected error index is used as the discriminator in the model.

3.2 Privacy-preserving back-propagation neural network-based learning algorithm

This paper proposes an algorithm for privacy-preserving back-propagation neural network-based learning in the case where the data set is horizontally partitioned.

When the data is horizontally partitioned, each participant owns all attribute values of some records or possesses some rows of records of the entire database. After the algorithm is executed, all participants can securely share the constructed learning model, and all participants can be used to predict the corresponding output for their target data.

Suppose the training data set D is horizontally partitioned into numbers; these n parts of the data set D after being partitioned are each owned by the participant $(P_1, P_2, \ldots P_n)$, where $|D_i| = n_i$. Each element $d_{i,j} \epsilon D_i, 1 \le j \le n$ is represented as shown in Equation (5).

$$d_{i,j} =< E_{i,j}. C_{i,j} >, E_{i,j} = \\ < 1, u_{i,j,1}, u_{i,j,2}, \ldots, u_{i,j,p} > \quad (5)$$

Where is the input vector, and is the corresponding output vector? The weight vector representation between each layer of neurons is shown in Equation (6).

$$W =< W_{p+1,0} \ldots, W_{p+1,p} \ldots, W_{p+k,0} \ldots, W_{p+k,p} \ldots, W_{p+k+1,p+1} \ldots, W_{p+k+1,p+k} \quad (6)$$

The purpose of neural network learning is to calculate the network weight vector W for the set of training samples. To protect the private information in the backpropagation neural network learning model from disclosure and to protect the knowledge of the weight vector W from exposure, the weight vector W can be assigned to all participants such that each participant owns a portion of the private value of the weight vector W. Let the weight vector owned by any of the participants be denoted as W_i. This is calculated as shown in Equation (7).

$$W_i =< W_{i,p+1,0} \cdots, W_{i,p+1,p} \cdots, W_{i,p+k,p} \cdots, W_{i,p+k+1,p+1} \cdots, W_{i,p+k+1,p+k} \quad (7)$$

where each element in the weight vector W can be obtained by $w_{p,s} = \sum_{i=1}^{n} w_{i,p,s}$.

Figure 2 shows a schematic diagram of the information propagation vector.

3.3 Correction dissemination

Here, we use a multidimensional interest model to analyze news propagation in the Leader-Follower network, and

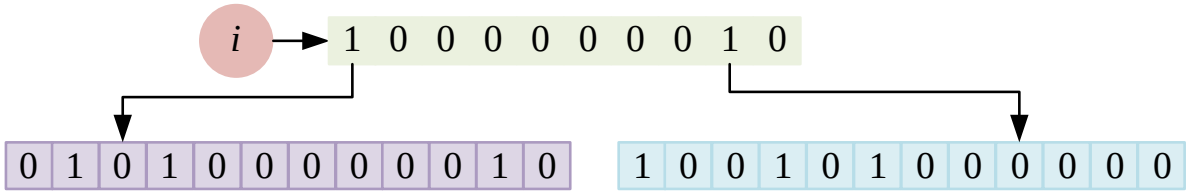

Figure 2: Schematic diagram of information dissemination. Indicates that the current user is interested in the 1-st major category and the 9th major category of the received i news information, interested in the 2nd, 4th, and 11th subcategories under the 1-st major type, and interested in the 1st, 4th, and 6th subcategories under the 2nd major category.

The current user satisfies M for the number of broad categories of interest in the ith news message received is the total number of broad categories of interest in the system. The limit on the number of user interests is to avoid users holding the same interest in all broad categories. The total number of elements with value 1 in the D-dimensional subcategory interest vector under the broad category represents the interest in one subcategory. For example, in Figure 2, The individuality of users in the multidimensional model is reflected in the differences in the major categories of interest of attention and the differences in the subcategory interest vectors of different users under the same major category, so this paper stipulates that the subcategory interest vectors t_i^c and t_j^c of any user i and j under the cth major category are not the same.

In order to elaborate in detail on how to apply backpropagation neural network learning algorithms to protect privacy in the case of dataset horizontal segmentation, we will provide a more in-depth description of the algorithm, including pseudocode and related security considerations. The training dataset D is horizontally divided into n parts, with each part D_i ($1 \leqslant i \leqslant n$) owned by participant P_i, where | D_i |=n_i. The weight vector W is initialized as a common random value or initialized according to a certain strategy. Determine encryption algorithms, security protocols, and possible thresholds and other related parameters. Each participant P_i calculates a certain summary of their local error (such as the average error) and sends this summary to a central coordinator (or uses a secure multi-party computation protocol for aggregation).

The central coordinator (or using secure multi-party computation protocols) calculates the direction and size of weight updates based on aggregated errors. This calculation can be based on the standard backpropagation formula, but it needs to be performed without exposing the original data. The updated weights are assigned (or incrementally updated) to each participant P_i. This can be achieved through various methods, such as using homomorphic encryption or secret sharing techniques.

3.4 Analysis methods and indicators

To verify that the adaptive news recommendation model has high user similarity and a high recommendation success rate. The average differences and average fraction metrics are used.

(1) Average Differences: Measures the degree of match between the interests of top and bottom users. The average interest difference between a user and his superior user is measured. The calculation is shown in Equation (8).

$$AD = \frac{1}{LU}\sum_{i\in U}\sum_{l\in L_i}(\sum_{c\in C_l}\|t_i^c - t_l^c\|/m_l \quad (8)$$

(2) Average Fraction: Measures user satisfaction with recommended news. The proportion of liked information among all that has been read is used as a measure. The calculation is shown in equation (9).

$$AF = \sum_{ia}(\delta e_{ia}, 1)/\sum_{ia}(\delta|e_{ia}|, 1) \quad (9)$$

Community partitioning is important to clarify the structure and function of the network. In this paper, we choose the gravitational-repulsive algorithm to partition the community of users in the Leader-Follower network. To facilitate further analysis and discussion of the community structure, the centroid in a community is defined as the user who has the most followers in the community and whose followers belonging to that community account for the largest proportion of all their followers.

The characteristics of news dissemination are reflected in the depth and breadth of dissemination; depth refers to the number of user layers influenced in the Leader-Follower network; Breadth refers to the extent of dissemination in the user network, i.e., the total number of users affected. In this paper, two propagation depths are defined based on the above definition: mini-depth, which means that for the same news received from multiple superiors, the depth is updated to the length of the path with the shortest news propagation path + 1; max-depth, which means that for the same news received from multiple superiors, the depth is updated to the length of the path with the longest news propagation path + 1.

For example, In Figure 3, the news is transmitted from user *I* to *J*. The minimum propagation path is *i*→*j*, the minimum propagation depth is 2; the maximum propagation path is i→x→j, and the maximum is 3.

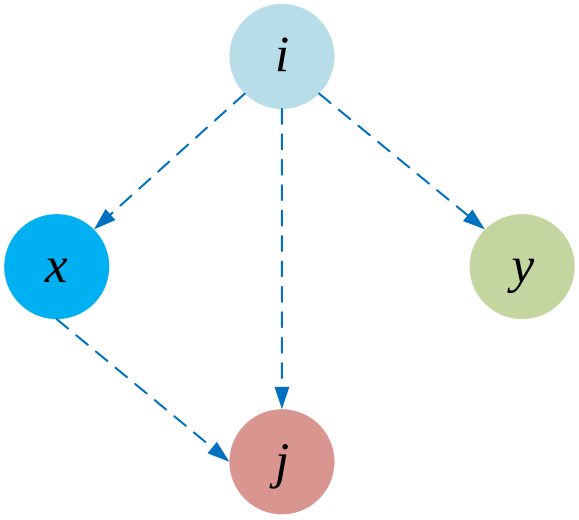

Figure 3: Diagram of news dissemination path

4 Experimental evaluation and analysis

4.1 Evaluation index and experimental environment

In this paper, a communication network based on microblogging data is used as the experimental social network, and three indicators are used for evaluation:

(1) Impact indicators: For the seed set S generated by the algorithm, the mean value of the influence IG(S) when the seed set S satisfies the privacy-preserving constraints is used as the evaluation metric, called the influence metric. The larger the influence metric is, the greater the influence generated by the seed set generated by the algorithm in the dissemination process and the better the results.

(2) Composite index: Considering that there are two objectives of the algorithm, maximizing the impact generated by propagation and satisfying the restrictions under the privacy protection constraint, the function *F* (G, S) is defined here as the evaluation metric of the algorithm, which is abbreviated as F(S) in the following and called the composite metric. The calculation is shown in Equation (10).

$$F(S)=\begin{cases}0, if\ \exists o_j \in O, Q(S,o_j)>\tau_j \\ I_G(S), otherwise\end{cases} \tag{10}$$

That is, F(S) is 0 when the seed set S does not satisfy the privacy-preserving constraints, and F(S) is equal to the size of influence IG(S) generated by the seed set S on the network in other cases. For the composite metric, the larger the metric is when the probability that the set S of seeds generated by the algorithm satisfies the privacy-preserving constraints is greater, the more effective the algorithm is. In addition to this, the metric increases when the seed set S produced by the algorithm generates more influence. Therefore, F(S) carefully considers the algorithm's privacy-preserving effect and the size of its propagation impact.

(3) Runtime metrics: For algorithms with similar effects, shorter running times imply better efficiency, and each algorithm’s running times are experimentally counted to evaluate their efficiency.

The experiments use Pytorch deep learning framework to build the model, the development language is Python, and the V100 32G GPU, Adam optimizer to optimize the model parameters. The model has an initial learning rate of 0.0001, a Batch Size of 32, a weight decay rate of 0.0005, and a momentum of 0.9. In addition, to prevent the model from overfitting, Dropout is set to 0.5. The training and testing Loss curves are shown in Figure 4.

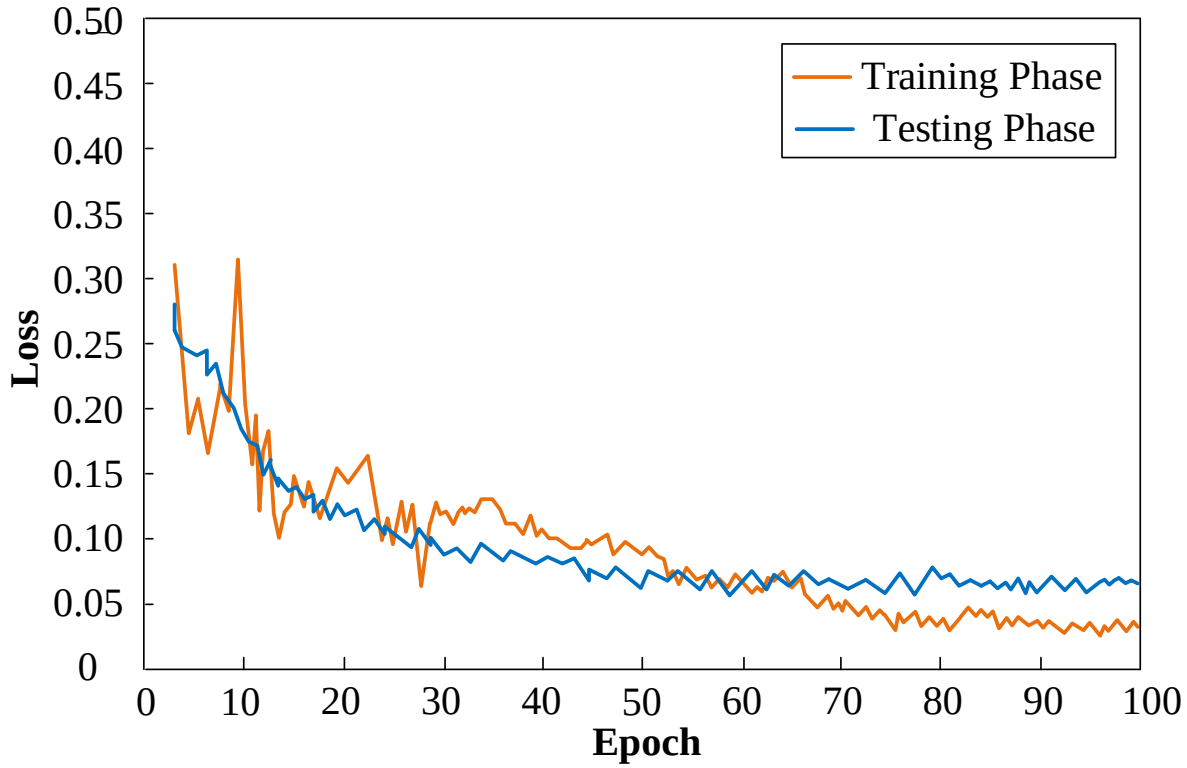

Figure 4: Loss curve

4.2 Analysis of results

(1) Execution efficiency comparison and analysis

The backpropagation neural network learning algorithm proposed in the literature [24] and the privacy-preserving backpropagation neural network learning algorithm proposed in this paper are selected for comparison regarding execution time. The experimental results are shown in Figure 5. With constant encryption key length and several participants, when the number of computational nodes is less than 42, the execution time required by the privacy-preserving backpropagation neural network learning algorithm in this paper is significantly lower than that of the comparison algorithm, verifying the efficiency of the method in terms of time overhead.

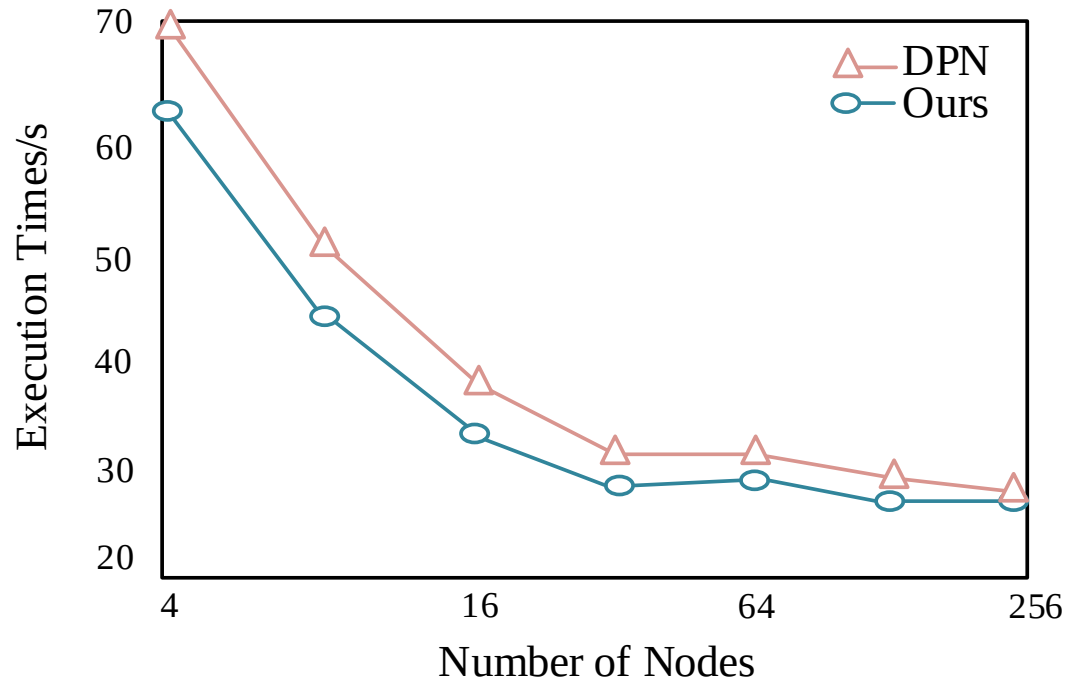

Figure 5: Time comparison

(2) Information dissemination success rate

In this paper, experiments are conducted separately for the model without the introduction of quality and for the model with the introduction of quality in the news release, with the same values of model parameters as in the literature [25]. The changes in Average Differences and Average Fractions with increasing time were analyzed according to the experimental results, as shown in Figure 6. As seen from Figure 6, the Average Differences of the multidimensional interest vector model with the introduction of quality converge to 4.6 as time increases, indicating an average of 4.6 subcategories of interest differences between adjacent users. In comparison, there are an average of 5.2 subcategories of interest differences between adjacent users in the model without the introduction of quality, which is less effective than the model with the introduction of quality. Average Fraction indicators are similar, converging to 0.65, indicating a recommendation success rate of 0.65, but the model with the introduction of mass combines faster.

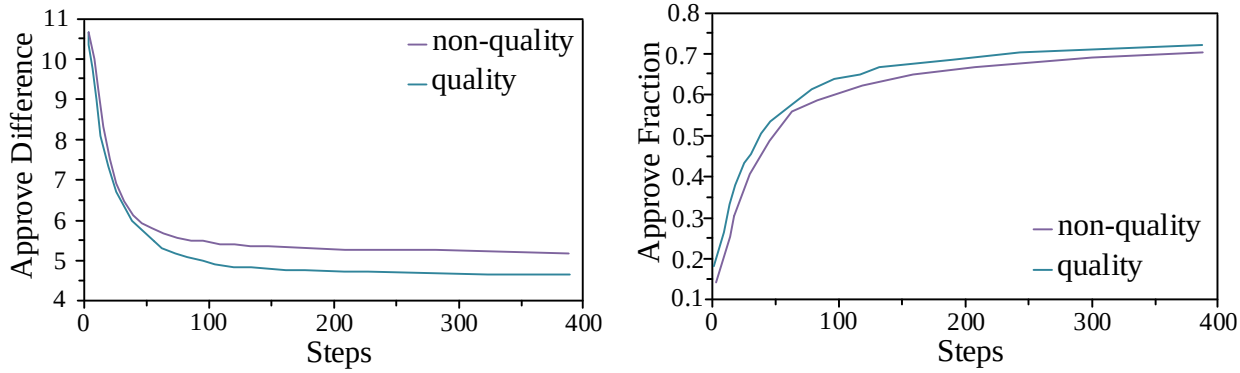

Figure 6: Similarity and recommendation success rate

(3) Model Privacy Rating Accuracy Test

In this paper, 1000 records are randomly selected from the training dataset as test samples, and the proposed privacy data metrics and grading model are tested for grading accuracy as follows. The test results are shown in Table 1.

Table 1: Model privacy classification accuracy test

Privacy Level	Sample size	Mistakes	Misjudgment rate (%)	Miscalculation deviation (%)
1	96	1	1.03	2.80
2	78	1	1.26	3.44
3	115	1	0.83	2.27
4	181	7	3.82	12.95
5	166	6	3.64	9.88
6	98	4	3.51	9.53
7	151	1	1.02	2.77
8	133	1	0.69	1.89
1~8	1000	22	2.20	6.45

The test results show that this paper's proposed privacy data metric and grading model can achieve an overall grading accuracy of 97.8% and a grading accuracy of over 96% for a single privacy-level sample. On the other hand, the misclassification rate of each privacy level shows that the model is more accurate in grading data at the relative privacy boundary levels (privacy levels 0, 1, 2, 6, 7) than those at the intermediate privacy levels (privacy levels 4, 6, 7), which is consistent with the reality that it is easier to measure the privacy of data when they provide more differentiated privacy information. This is also reflected in the misclassification rate of each confidentiality level, and the trends of misclassification rate (E) and misclassification bias rate for each group are shown in Figure 7 and Figure 8.

As seen from Figure 7, the trend of the degree of misclassification bias of the model for the dataset as a whole, i.e., the misclassification bias rate, is similar to the misclassification rate. It is related to $\left|D_i - D_i^{'}\right|$ a power series, so it should also be connected to a power series, and Figure 8 shows that the variation tends to be a straight line near the natural logarithm e. Therefore, it can be inferred that the values are mostly 0 and 1, which indicates that the model rarely misclassifies across levels when misclassification occurs, i.e., the results in which misclassification occurs are basically in the adjacent privacy levels. Such a degree of misclassification bias in the model is acceptable when the misclassification rate is not high.

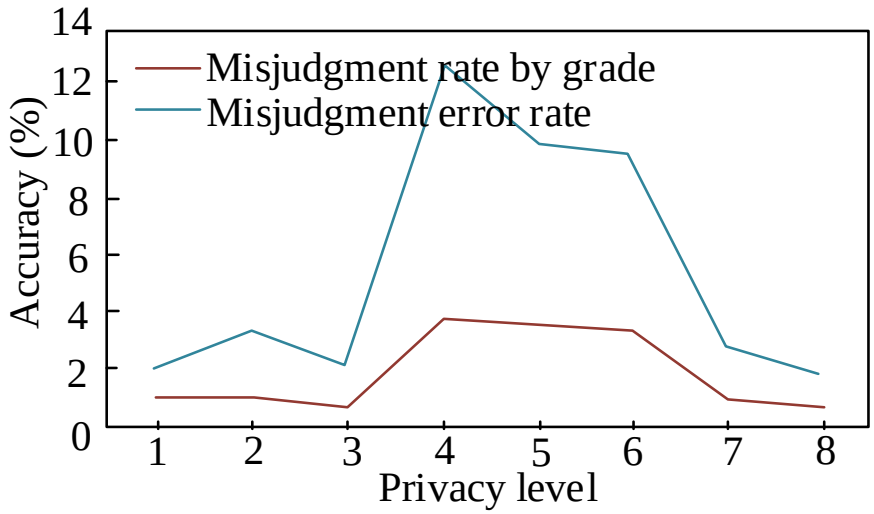

Figure 7: Change curve of misjudgment rate and misjudgment error rate

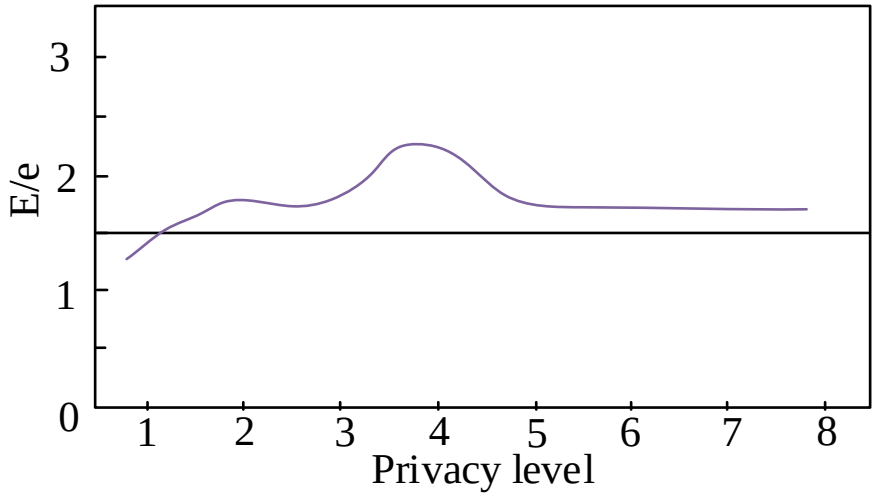

Figure 8: Change curve of E/e

(4) Depth, breadth, and acceptance of news dissemination

Figure 9 portrays the distribution of the minimum depth (min-depth) and maximum depth (max-depth) of news dissemination without introducing quality and with introducing quality factors in news releases, respectively. From Figure 9, we can see that most news has a shallow depth of dissemination, only 0 to 2 steps, while the proportion of information with a very deep depth of dissemination is very small. When the quality factor is not introduced, there are two peaks of news dissemination depth, the first peak represents more news dissemination depth within 2, and the second peak represents more news dissemination depth moderate, representing a category of news that is not particularly good (mediocre). Introducing a quality factor to news releases resulted in more information appearing within the first peak. In contrast, the second peak was flatter than the data without introducing quality, and the depth of dissemination of good news increased. This suggests that the introduction of quality makes the recommendation of information more differentiated, while the amount of data disseminated in the network decreases as the number of mediocre news decreases, but does not inhibit the dissemination of good reports; rather, good news spreads more deeply.

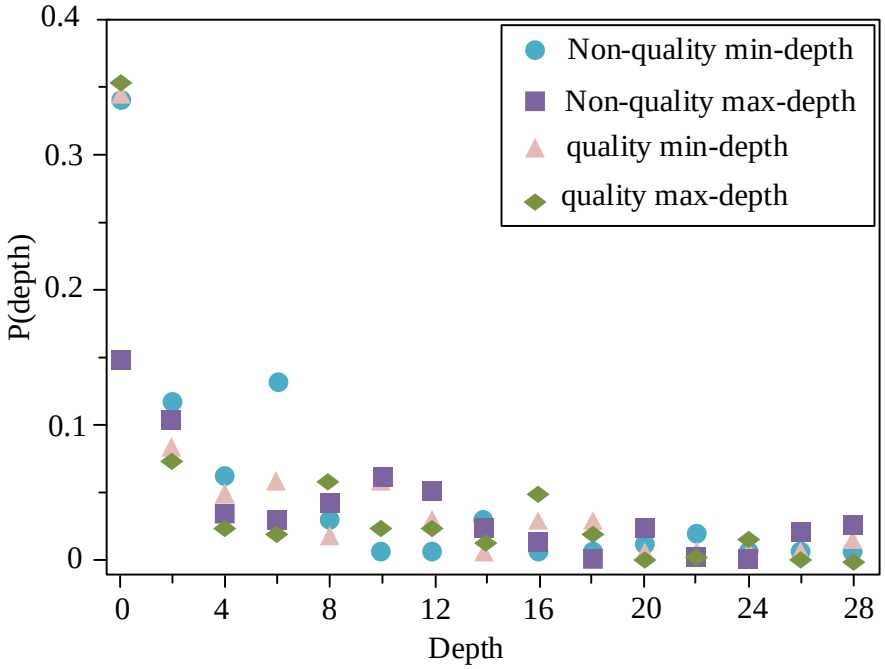

Figure 9: Effect of introduction quality on the minimum-maximum depth of news dissemination

The relationship between the maximum depth (max-depth) and standard deviation (SD-depth) of news dissemination and news quality after the introduction of quality is shown in Figure 10.

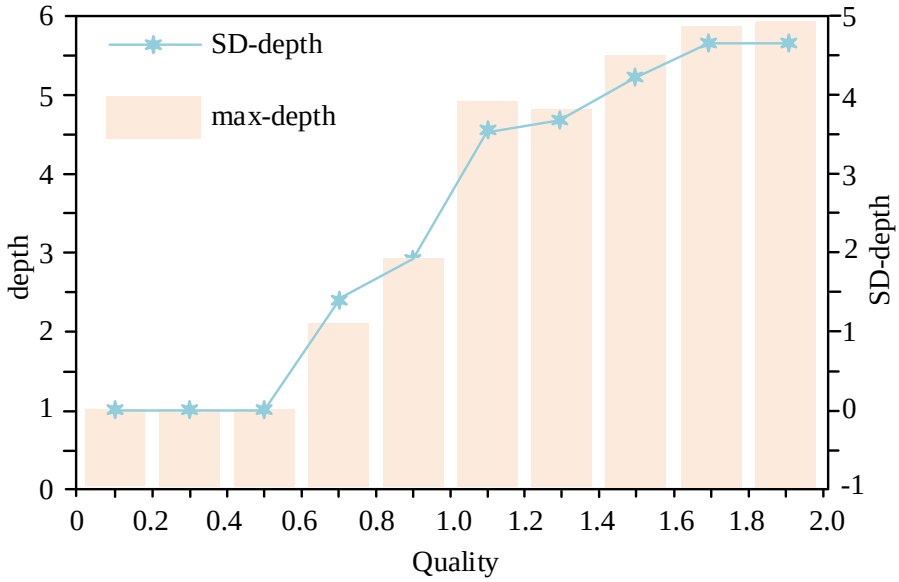

Figure 10: Relationship between maximum depth and standard deviation of news dissemination and quality

As shown in Figure 9, the depth of news dissemination increases as quality increases and the standard variance widens. It means that as the quality increases, the news belonging to high quality not only increases the depth of dissemination but also the gap in the depth of dissemination, and there is a greater possibility of some far-reaching news among the high-quality news. As for some low-quality information, most have very little depth of dissemination, reflected in a small average depth and a small standard deviation.

5 Discussion

In this article, we propose a new news dissemination information model based on BP neural network, combined with user privacy protection methods. Through comparison and analysis, we found that the model exhibits significant performance and advantages in multiple aspects. Firstly, from the perspective of privacy protection, our proposed privacy data measurement and rating model achieved an accuracy of up to 97.8% in privacy level classification, which exceeds the performance reported in previous studies. It is particularly noteworthy that our model shows higher accuracy in data classification at the

relative privacy boundary level, which is consistent with the actual situation, that is, when the data provides more differential privacy information, the accuracy of measuring data privacy will be higher. The improvement in accuracy is mainly due to the scoring model we designed, which can more finely capture the subtle differences between different privacy levels.

Secondly, from the perspective of news communication, our model significantly improves the maximum depth and standard deviation of news communication by introducing quality factors. This indicates that high-quality news is not only more easily disseminated in depth, but also has greater differences in depth of dissemination, thereby increasing the possibility of producing news with far-reaching impact. This discovery has important practical significance for news recommendation and social media management, as it can help us better understand and optimize news dissemination strategies.

Compared to traditional and modern models, our model demonstrates advantages in multiple aspects. Firstly, in terms of privacy protection, we adopted a privacy data measurement and rating model based on BP neural network. This model can not only evaluate the privacy level of data without leaking the original data, but also more accurately classify data with different privacy levels. Secondly, in terms of news dissemination, we have introduced quality factors to optimize the dissemination strategy of news, thereby improving the effectiveness and efficiency of news dissemination. Finally, in terms of execution time and error measurement, our model is able to complete training and prediction tasks in a relatively short amount of time, and has a low error classification rate and bias rate.

6 Conclusion

This paper proposes a new BP neural network-based news dissemination information model with a user privacy protection method. The algorithm is suitable for distributed environments where multiple participants exist, where all participants jointly construct a neural network learning model over the entire data set, and where each participant does not need to reveal their data to the other participants. Finally, all participants can securely share the learning model and use it to predict the appropriate output for their target data. Future work will discuss how to design privacy-preserving backpropagation neural network-based learning algorithms when the data is vertically segmented.

Authorship contribution statement

Jingjing Guo: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

Jianqiang Wang: Methodology, Software, Validation.

Data availability

On Request

Declarations

Not applicable

Competing of interests

The authors declare no competing of interests.

References

[1] P. Zhang, Y. Wang, N. Kumar, C. Jiang, and G. Shi, "A security-and privacy-preserving approach based on data disturbance for collaborative edge computing in social IoT systems," *IEEE Trans Comput Soc Syst*, vol. 9, no. 1, pp. 97–108, 2021.

[2] S. Ghimire, Z. M. Yaseen, A. A. Farooque, R. C. Deo, J. Zhang, and X. Tao, "Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks," *Sci Rep*, vol. 11, no. 1, p. 17497, 2021.

[3] G. Li, "The Influence of News Dissemination in the Micromedia Era Based on the Deep Trust Network Model," *Mobile Information Systems*, vol. 2022, 2022.

[4] M. Wang, "Artificial Intelligence-Driven Model for Production Innovation of Sports News Dissemination," *Wirel Commun Mob Comput*, vol. 2022, 2022.

[5] X. Zhou, A. Jain, V. V Phoha, and R. Zafarani, "Fake news early detection: A theory-driven model," *Digital Threats: Research and Practice*, vol. 1, no. 2, pp. 1–25, 2020.

[6] Y. Ge, "The influence of news communication for production mode on computer news dissemination [J]," *The Frontiers of Science and Technology*, vol. 3, no. 04, pp. 14–17, 2021.

[7] R. F. Muhammad and S. Kasahara, "A trust model for information dissemination in social networking services," *IEICE Proceedings Series*, vol. 63, no. D3-1, 2020.

[8] N. K. Agarwal and F. Alsaeedi, "Creation, dissemination and mitigation: toward a disinformation behavior framework and model," *Aslib Journal of Information Management*, vol. 73, no. 5, pp. 639–658, 2021.

[9] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional LSTM-recurrent neural network," *Procedia Comput Sci*, vol. 165, pp. 74–82, 2019.

[10] A. Aggarwal, A. Chauhan, D. Kumar, S. Verma, and M. Mittal, "Classification of fake news by fine-tuning deep bidirectional transformers-based language model," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 7, no. 27, pp. e10–e10, 2020.

[11] T. A. Maniou and A. Veglis, "Employing a chatbot for news dissemination during crisis: Design, implementation and evaluation," *Future Internet*, vol. 12, no. 7, p. 109, 2020.

[12] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science (1979)*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[13] P. M. Waszak, W. Kasprzycka-Waszak, and A. Kubanek, "The spread of medical fake news in social media–the pilot quantitative study," *Health Policy Technol*, vol. 7, no. 2, pp. 115–118, 2018.

[14] T. Murayama, S. Wakamiya, E. Aramaki, and R. Kobayashi, "Modeling the spread of fake news on Twitter," *PLoS One*, vol. 16, no. 4, p. e0250419, 2021.

[15] Q. Feng, D. He, S. Zeadally, M. K. Khan, and N. Kumar, "A survey on privacy protection in blockchain system," *Journal of network and computer applications*, vol. 126, pp. 45–58, 2019.

[16] P. Sun, "Security and privacy protection in cloud computing: Discussions and challenges," *Journal of Network and Computer Applications*, vol. 160, p. 102642, 2020.

[17] P. J. Sun, "Privacy protection and data security in cloud computing: a survey, challenges, and solutions," *IEEE Access*, vol. 7, pp. 147420–147452, 2019.

[18] D. Zhang, "Big data security and privacy protection," in *8th international conference on management and computer science (ICMCS 2018)*, Atlantis Press, 2018, pp. 275–278.

[19] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in transformer," *Adv Neural Inf Process Syst*, vol. 34, pp. 15908–15919, 2021.

[20] U. M. Rao, I. Fofana, T. Jaya, E. M. Rodriguez-Celis, J. Jalbert, and P. Picher, "Alternative dielectric fluids for transformer insulation system: Progress, challenges, and future prospects," *IEEE Access*, vol. 7, pp. 184552–184571, 2019.

[21] Q. Liu *et al.*, "Comparative analysis of BP neural network and RBF neural network in seismic performance evaluation of pier columns," *Mech Syst Signal Process*, vol. 141, p. 106707, 2020.

[22] J. Li, X. Yao, X. Wang, Q. Yu, and Y. Zhang, "Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis," *Measurement*, vol. 153, p. 107419, 2020.

[23] W. Wang, R. Tang, C. Li, P. Liu, and L. Luo, "A BP neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights," *Ocean Engineering*, vol. 162, pp. 98–107, 2018.

[24] J. Xiong *et al.*, "A personalized privacy protection framework for mobile crowdsensing in IIoT," *IEEE Trans Industr Inform*, vol. 16, no. 6, pp. 4231–4241, 2019.

[25] X. Mengyao and W. Qian, "Analysis of news transmission mode based on fuzzy data classification and neural network simulation," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 6, pp. 7133–7143, 2020.