Mechatronic Systems and Control, Vol. 51, No. 10, 2023
An Open Access Paper
Special Issue
Neural Networks and Fuzzy Logic in Mechatronics and Robotics

INTERNET INFORMATION COLLECTION AND DATA ANALYSIS BASED ON ARTIFICIAL INTELLIGENCE

Yongtao Zhang* and Na Yang*

Abstract

With the accelerated advancement of computer and network communication technology, the explosive growth of Internet information, the advancement of science and technology, and the advancement of Internet technology, contemporary artificial intelligence is also advancing quickly, and the ongoing advancement of machine learning is accompanied by a significant number of new developments. And the amount of jumbled information data has significantly expanded. Today, practically every user must search through a lot of stuff online every day. We need to apply artificial intelligence to better gather, organise, and store these vast volumes of information data. By gathering large amounts of data and conducting simulation tests, it is possible to efficiently address the issues of a lengthy information collection process and low analytical data accuracy. Insufficient important information will be produced by data collection and processing, making it impossible to satisfy users in terms of accuracy and applicability. However, the two well-known data systems Hadoop and Spark handle comparable issues and enable quick and precise analysis of large amounts of data in the context of modern big data.

Key Words

Artificial intelligence, information collection, data analysis, convolutional neural network

1. Introduction

With the development of modern artificial intelligence, in order to better meet the processing needs of the collection and analysis of Internet information, the convolutional neural network model can be optimised according to the development characteristics of artificial intelligence [1]–[3] and used in combination with the metric multi-dimensional network model algorithm [4]–[6] to improve efficiency in actual computing and application. In order to reduce the false alarm rate of artificial intelligence in Internet information collection and analysis and improve the accuracy rate of information collection, a measurement multi-dimensional network model and Trace-DF convolutional neural network technology that can accurately collect Internet information and analyse data are used [7]–[9]. Create a smart system with the intention of using it for data analysis and information gathering on the Internet. Results from practise have shown to increase productivity. Recognising the benefits of the convolutional neural network model and paying attention to its network model are necessary to accomplish this goal while also increasing the effectiveness of information gathering. However, due to the complexity of the network information situation, this could result in the collection of ambiguous data that cannot be analysed. The final test demonstrates that the processing effect of entropy-weighted data model programming has been greatly enhanced, boosting the data gathering and data analysis capabilities of artificial intelligence on the Internet [10], [11]. With the development of science and technology, in view of the intelligent needs of its model, a model using relevant data for information preprocessing was designed, and experiments showed that in the work operation, the MapReduce analysis model can ensure the accuracy and stability of its information collection [12], [13]. With the rapid development of modern science and technology and the advent of the network era, automation technology is more and more widely used in information collection and data analysis, with the characteristics of automation and intelligence, and its application plays an important role in information collection and data analysis [14]. It can not only satisfy the information collection's accuracy standards but also enhance data analysis skills and offer a solid assurance for users' desire for information on the Internet. The total quality of information data collection and analysis is increased by an understanding and analysis of artificial intelligence in Internet data collecting and analysis [15].

2. Overall Design of Internet Information Collection and Data Analysis Methods and Steps

After decades of continuous development, artificial intelligence and artificial neural networks for the development of

* Department of Information Engineering, Shangqiu Institute of Technology, Shangqiu 476000, China; e-mail: ytzjw@sqgxy.edu.cn; 15709744475@163.com; 1350003009@sqgxy.edu.cn
Corresponding author: Na Yang

Recommended by Farbod Khoshnoud
(DOI: 10.2316/J.2023.201-0348)

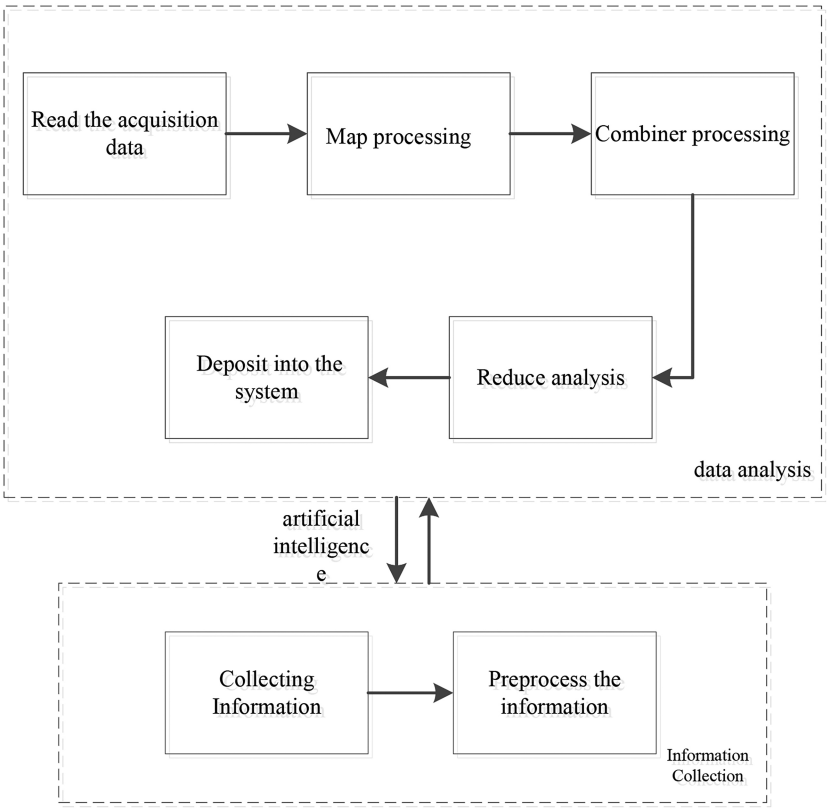


Figure 1. Acquisition and analysis flowchart.

Internet network information collection and data analysis technology have made an important contribution, the main idea of information collection and data analysis is to find a similar data model in the network database, and finally complete the relevant data analysis and processing through the identified data, in the information collection, the most common methods are artificial nervous system method, measurement dimension method, based on accounting method, etc., on the Internet, we use feature recognition to control network collection of information, It has greater advantages for Internet information collection and data analysis. With the rapid development of modern science, all things can be digitally presented to the public on the Internet, so how to collect and analyse the required information on the Internet is what current artificial intelligence needs to learn. In the big data environment, the obtained relevant raw data cannot be used directly, and it is difficult to give full authority to the terminal equipment to judge subjectively. Therefore, in the face of a large amount of useless data on the Internet, the information is collected and analysed with the help of artificial intelligence, and a simple classification judgement is first made on the results of the collection and arrangement, so as to complete the work of target information collection and data analysis. The acquisition and analysis flowchart is shown in Fig. 1.

Map processing is a method used for array data, which can perform "callback" for each collected data and return the analysis results as new data. In other words, by collecting a process to be performed in this analysis flow, a relevant analysis operation can be performed on each data of the relevant Internet information.

As shown in Fig. 1, in the process of information collection and data analysis and processing on the Internet, we first collect the required information, then preprocess the collected information, then collect the data through artificial intelligence, and then analyse it through MapReduce. The corresponding information is obtained, and then the data is transmitted to the system through artificial intelligence, and the system processes the operation results and submits the data to the user [16]. In the process of Internet information collection and data analysis, the strong correlation between information collection and data analysis is combined with the assistance of artificial intelligence. The information collection and analysis system is fast and accurate. Data collection and analysis require some new network models to calculate, and the data after the collection of Internet information is required on the premise that the results are correct. The optimisation of algorithms under big data technology and more scientific improvements are also worth looking forward to and paying attention to in the future.

MapReduce is a high-performance parallel computing platform based on cluster. MapReduce analysis provides us with a huge and sophisticated data analysis framework, which can automatically complete the parallel processing of Internet data collection and analysis, automatically calculate and analyse data, automatically distribute relevant data information on the collected data nodes and collect analysis results. Store data in a distributed manner. Spark framework is a kind of object-oriented and functional programming, which can operate Internet data collection and realise the ability to process massive data.

2.1 Overview of the Information Collection and Data Analysis Process

Currently, gathering general information online is similar to improving pertinent data analysis techniques. Although this strategy can make information collecting more effective, it is unable to address the issue of worthless data in the information gathered through analysis. When there is a lot of data being collected, there will also be more useless data, which will affect how usable data is analysed and extracted. Therefore, an automatic encoder is used to suggest a way for gathering information in order to address the issue of garbage information in information collection. A multidimensional autoencoder model for data processing and information gathering.

2.2 Measure the Multi-dimensional Network Model

Although the general neural network model can achieve good results for simple problems, but with the increasing complexity and difficulty of the problems required by users, the simple model has not been enough to achieve satisfactory results, so it is necessary to refer to the multi-dimensional automatic coding network model improved by measuring the learning thinking to preprocess the network data, and enhance the network information data by measuring the multi-dimensional network model, and reduce the irrelevant information in the collection features. Using the advantages of this model in collecting information and extracting information data, it can better deal with the problem of information collection on the Internet, and introduce a measurement model to form a measurement multi-dimensional network model [17], [18].

2.2.1 Metric Multi-dimensional Network Model Design

The model mainly aggregates and organises the network information of the same category by analysing the similarity degree and measures the input data of the multi-dimensional network model. The $x = \{x_1, x_2, x_3, \ldots, x_N\}$ encoding process is as follows:

$$a^n = f(x^n) = l(Wx^n + b) + d(Wx + b) \tag{1}$$

W and b represent the weights and biases of the a^n encoding part, respectively, and represent the resulting feature map, which is then decoded:

$$\overset{\Lambda}{a^n} = g(a^n) = l(Wa^n + b) - d(Wx + b) \tag{2}$$

In order to achieve information collection, the original information is mapped to the feature space, and the extracted information features are analysed. In the analysis process, multiple data acquisition may be used, which leads to the phenomenon of repeated data acquisition. In order to prevent duplicate data acquisition, the metric setting is used in this paper. A distance parameter d is defined.

$$d = \delta\left(s_i + \sum_j a^n y_j^{(h)} + \sum_j a^n y_j^{(l)}\right) \tag{3}$$

By continuously adjusting the parameter size of d, setting an appropriate collection formula, and calculating through the measurement model, the target information is screened multiple times through a nonlinear transformation formula. According to this idea, the objective function of measuring the multi-dimensional network model is:

$$l(\theta) = (1/2n)\sum_{j=1}^{n}(\lambda||h^{(j)} - h^{(j)}||_1 + |\eta|\overline{x}^{(j)} - x^{(j)}||_2^2) \tag{4}$$

The feature formula learned by the metric multi-dimensional network model is:

$$h = \delta(Wx + b) + \delta(Wd + b) \tag{5}$$

Two one-layer neural networks are superposed together to form a two-layer neural network, and the collected information is analysed and constructed for the second time to ensure that the screened data reduces the duplication problem.

The measurement multi-dimensional network model function's first half constructs the information that has been collected, and the second uses nonlinear transformation to reduce the distance between similar samples, minimise the impact of outside influences, and maximise the retention of the primary information that has been collected. The function for information loss is defined:

$$La = yd_{\mathrm{la,lb}}^2 + (1-y)(a - d_{\mathrm{la,lb}})^2 + (1-y) \tag{6}$$

$$A = (b_{\mathrm{ij}})_{m\times n} = (w_j \times a_{\mathrm{ij}})_{m\times n} \tag{7}$$

When the collected information increases, it $d_{\mathrm{la,lb}}$ will gradually increase to meet the need for information collection.

2.3 Entropy Weight Data Decision Analysis Model

When encountering less sample data, the network model cannot obtain enough information from the data, the neural network lacks the problem of underfitting, in order to allow the neural network model to collect Internet information from multiple angles and analyse related data, 10,000 square meters proposed an entropy decision analysis model, the model can collect feature information through multiple angles, and the feature information obtained by the operation will become the key point for the neural network to collect information. According to the entropy theory, the entropy weight data model is a calculation method [19], which is used to express the amount of information on related events. The smaller the amount of information, the greater the uncertainty; conversely, the greater the amount of information, the greater the uncertainty. The smaller the entropy, the smaller the entropy. Its calculation process is as follows:

Assuming that the forward processing is X_{ij}, the calculation proportion is P_{ij}:

$$P_{\mathrm{ij}} = \frac{x_{\mathrm{ij}}}{\sum_{i=1}^{m} x_{\mathrm{ij}}}(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n) \tag{8}$$

Entropy value e_i:

$$e_i = -k\sum_{i=1}^{m} p_{\mathrm{ij}} 1np_{\mathrm{ij}} (j = 1, 2, \ldots, n) k \geq 0, e_j \geq 0 \quad (9)$$

Differences g_i:

$$g_i = 1 - e_i (j = 1, 2, \ldots, n) \quad (10)$$

Weight w_j:

$$w_{\mathrm{ij}} = \frac{g_i}{\sum_{i=1}^{n} g_i} (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n) \quad (11)$$

Based on the above information, it can be concluded that the larger the factor value difference, the larger the weight value, and *vice versa*. In order to analyse the collected information, an ELECTRED algorithm [20], [21] that can compare the order relationship is used to sort and analyse the collected information and remove useless data. The algorithm steps are as follows:

$$W = (w_1, w_2, \ldots, w_n), w_i (i = 1, 2 \ldots, n)$$

Determine the weights to check, that is:

$$J^+(x_i, x_r) = \{j | 1 \leq j \leq n, x_{\mathrm{ij}} \rangle x_{\mathrm{rj}}\} \quad (12)$$

Similarly:

$$J^-(x_i, x_r) = \{j | 1 \leq j \leq n, x_{\mathrm{ij}} \langle x_{\mathrm{rj}}\} \quad (13)$$

Harmony Index:

$$1_{\mathrm{ir}} = \left(\sum_{j \in J^+(x_i, x_r)} w_j + \sum_{j \in J^-(x_i, x_{r)}} w_j \right) \Big/ \sum_{j=1}^{n} w_j \quad (14)$$

$$I'_{\mathrm{ir}} = \left(\sum_{j \in J^+(x_i, x_r)} w_j \right) \Big/ \left(\sum_{j \in J^-(x_i, x_r)} w_j \right) \quad (15)$$

The ELECTRED algorithm is used to detect relevant data, clarify the level relationship of the collected data, and then analyse and sort the target data [22], [23]. Once the data analysis is sorted, the amount of relevant calculations can be reduced.

3. Improvement of Internet Information Collection and Data Analysis Algorithms

After the Collection of Internet information, the collected information will be summarised to get a huge amount of information database, our database will often retain the invalid information we collected, and then the introduction of decision tree model can effectively solve the problem of data duplication. The Decision Tree network model is primarily a nonparametric data classifier with simple usage and low difficulty of operation. Here, we quote the FM algorithm, as well as the second-order term parameters for computation [24], [25].

FM is a data analysis predictor that can work with any real-valued feature vector. Other recent data analysis test models are only available for very limited input data. But we can let FM make the latest model measurement by defining the feature vector of the input data.

The commonly used algorithm in the decision tree network model is the FM algorithm, which can be used to divide and organise the collected data set, and the ultimate goal of the decision tree node splitting is to make the data samples on each collected information in the same category to the greatest extent, which is simply to collect data with higher accuracy. the gain of each feature information is calculated separately through the data model and, finally, the feature of the maximum information gain is selected to do the node splitting of feature acquisition, which further improves the accuracy of the decision tree for Internet information collection and analysis, and the collected information data is studied by the FM algorithm for correlation study, and the same data information is divided into the corresponding processing system, where the stronger the ability to collect information after processing, the stronger the characteristics of the information data.

According to the previous model and formula, we have carried out a preliminary architecture of the big data technology algorithm of Internet information collection and data analysis, but there are still deficiencies and loopholes, and we will optimise and improve our relevant network model algorithms below to improve the computing power and accuracy of the algorithm.

For the algorithm problems of Internet information collection and data analysis, in addition to the metric model algorithm, the FM algorithm can be used for subsequent data analysis and to process the relationship between two related data and analyse the correlation between the collected data. We deduce the correlation, assuming that there are two correlation functions f and g, then the correlation formula is described as:

$$(f \otimes g)(y) = \int_{\infty}^{-\infty} f*(t) g(t + y) \mathrm{dt} \quad (16)$$

$$y = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{\mathrm{ij}} x_i x_j \quad (17)$$

In the formula, n represents the number of features of the sample, and x represents the model parameters. The analysis of the model from the formula alone is to find the most responsive value of the relevant target in the data on the Internet. The simplest way to understand this formula is that the closer the two data values are the higher the correlation value.

The second-order parameter is used to reduce the model parameter error. The second-order algorithm processes the target through a relevant low-rank matrix. The relevant formula is:

$$w_{\mathrm{ij}} = {v_i}^T v_j, V \in R^{k*n} \quad (18)$$

For each data feature, an auxiliary vector (hidden vector) V is introduced, and then w_{ij} is solved using V_i.

Therefore, the algorithm first calculates the parameters of the matrix w, and then calculates the latent vector v. When calculating larger problems, the fast Fourier

transform operation is used to reduce the amount of related calculations and improve efficiency. The formula is as follows:

$$G = F \bullet H^* \tag{19}$$

G represents the response output, F represents the relevant data, $H*$ represents the filter, and the formula to be solved is

$$H^* = G/F \tag{20}$$

In order to ensure the accuracy in the calculation process of the model formula and consider the influence of related factors on the results, the FM regression algorithm is added to transform the formula, which reduces the computational complexity of the algorithm. The formula is:

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(v_i^T v_j)x_i x_j = \frac{1}{2}\sum_{i=1}^{k}(\left(\sum_{i=1}^{n} v_{ij}x_i\right)^2 - \sum_{i=1}^{n} v_{ij}^2 x_i^2 \tag{21}$$

Then the above formula is calculated by θ the gradient descent method:

$$\theta = \theta_0 - \eta * \nabla f(\theta_0) \tag{22}$$

For the length of each iteration η, in order to ensure that the current data is not far from the initial target demand data, the new adagrad algorithm formula is used:

$$A_t = \eta F_t \bullet G_t^* + (1-\eta)A_t - 1 \tag{23}$$

$$B_t = \eta F_t \bullet F_t^* + (1-\eta)B_t - 1 \tag{24}$$

$$\theta_{t+1,j} = \theta_{t,i} - \frac{\eta}{\sqrt{\sum_{i=0}^{1}(g^i)^2}} g' \tag{25}$$

Among them, A_t is the numerator of the current formula, B_t is the denominator of the current formula, and the data parameter θ.

Some data collection and applications are also implemented in wearable technology [26], and data collection has important application scenarios in intelligent algorithms. The use of artificial intelligence technology in data transmission can effectively improve the efficiency and economic value of data [27], [28].

4. Experimental Simulation

In order to ensure the compatibility of related models and algorithms for Internet information collection and data analysis, a large amount of data is used for experimental verification. The experimental objects are: the comparison of information collection efficiency, the success rate of data analysis, and the time-consuming for successful information collection. Compare and determine the advantages of the model algorithm before and after the improvement. The preprocessing of the collected information has been described in the preface. We then search for its keyword feature points, then collect and classify the collected information, analyse, and sort it out, check for errors, and output the data if there is no problem. To the user, try to improve the accuracy as much as possible, as shown in Fig. 2.

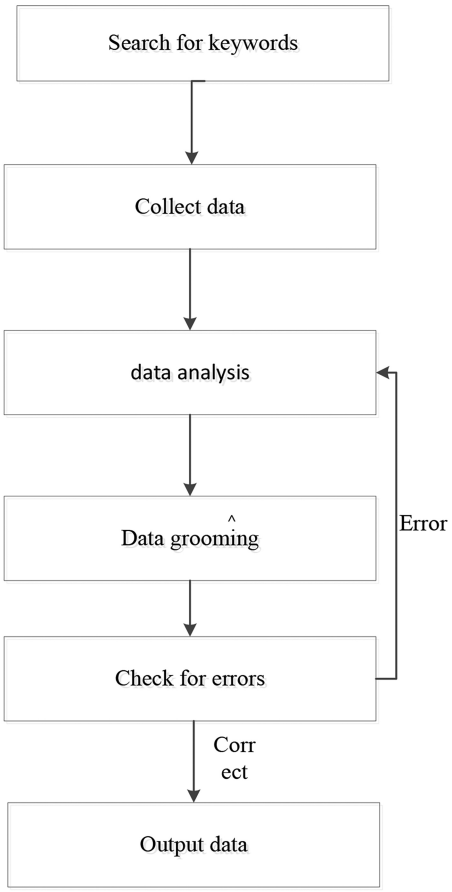


Figure 2. Collect and analyse information for decision making.

4.1 Comparison of Models Optimised Using Neural Networks

Compared with measuring multi-dimensional network models, using the optimised tracking algorithm network model data is more systematic and smoother. It removes two fully connected layers and adds two convolutional layers and a global filter model on the original basis, which greatly reduces the relevant parameters required by the tracking algorithm model, from 2.30 M to measure the multi-dimensional network model. The network parameters are reduced to 20 M parameters. The following is the information pre-collection data of the two models under the same network data set, as shown in Table 1 and Fig. 3.

By comparing the information collection, training, and testing of different levels of Internet data, the classification effects of the three scales of the Trace-DF model in the R SHOA-4 dataset are compared. According to these data, we can prove that the parameters, the Trace-DF convolutional neural network has a higher classification accuracy, which is enough to explain the improvement of

Table 1
Trace Algorithms Network Model in R SHOA-4 Data Set Information Acquisition Experimental Data Table

CNN	Train	Test	Iterations	Training Time	AA
Trace-ALGO (256×256)	480	240	3000	30 min	92.0
Trace-DF (512×512)	480	240	3000	20 min	92.6
Trace-DF (256×256)	480	240	3000	15 min	91.3
Trace-DF (128×128)	480	240	3000	12 min	90.8

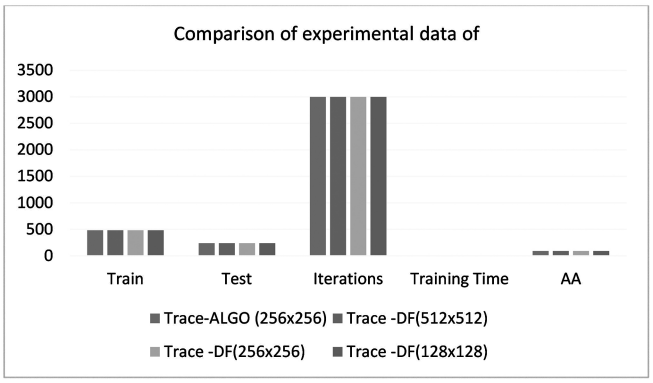


Figure 3. Comparison of experimental data of Trace-DF model under different data.

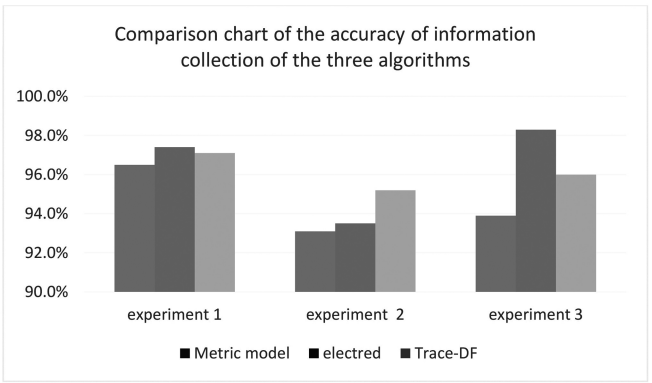


Figure 4. Comparison of information collection accuracy of three algorithms.

Table 2
The Experimental Data Table of the Information Collection Accuracy of the Three Algorithms

Algorithm	Experiment 1	Experiment 2	Experiment 3
Metric model	96.5%	93.1%	93.9%
ELECTRED	97.4%	93.5%	98.3%
Trace-DF	97.1%	95.2%	96.0%

the Trace-DF convolutional neural network for Internet information collection.

By comparing the Metric model algorithm, the ELECTRED algorithm and the Trace-DF algorithm on the Internet through the convolutional neural network on the R SROD-3 data set for target detection, the average accuracy of the three algorithms for information collection is shown in Table 2 and Fig. 4.

From the comparison chart and experimental data, it can be seen that the Metric model algorithm has the lowest average accuracy in the three scenarios, and the ELECTRED algorithm and the Trace-DF algorithm have higher accuracy in collecting information on the Internet. Through comparison, we can choose a model that meets our needs and put it in our Internet information collection and data analysis algorithms. Then through the experimental data to observe the average missed alarm rate data of the three algorithms for Internet information collection and data analysis under different information platforms, as shown in Table 3 and Fig. 5.

Table 3
Experimental Data Table of the Average Missed Alarm Rate of the Three Algorithms under Different Information Platforms

Algorithm	Amazon	Ebay	Microsoft
Metric model	47.5%	38.9%	21.1%
ELECTRED	24.6%	17.5%	10.3%
Trace-DF	6.9%	7.8%	10.5%

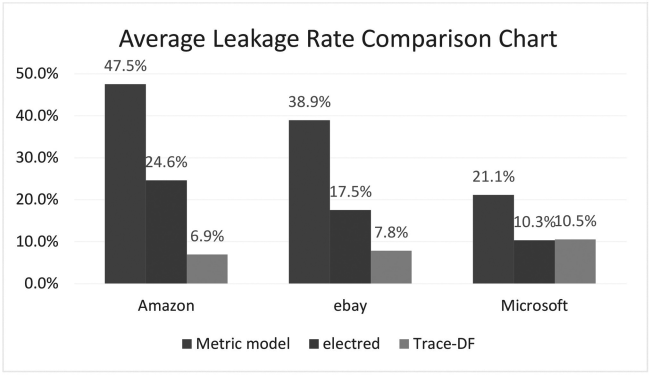


Figure 5. Comparison of the average false alarm rate of the three algorithms.

Through these two sets of experiments, we can confirm that the convolutional neural network can achieve better detection results in the experiments of information

Table 4
Time-consuming Comparison of Three Algorithms

Algorithm	Information Collection Time (ms)	Time Spent Analysing Data Results (ms)	Total Time (ms)
Metric model	602.6	625.5	1228.1
ELECTRED	447.9	420.2	868.1
Trace-DF	356.6	367.5	724.1

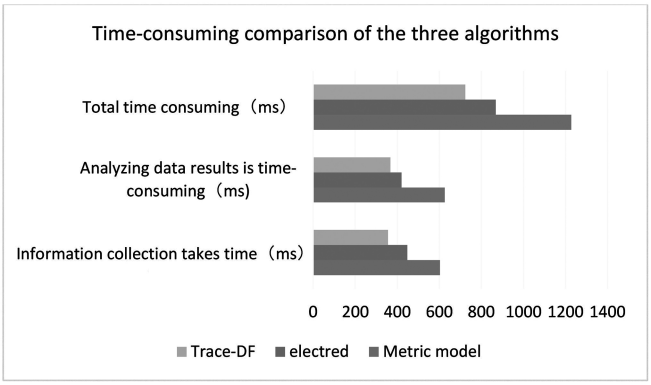


Figure 6. Time-consuming comparison of three algorithms.

collection and data analysis and detection on the Internet. Through the experimental data, it can be seen that the elected algorithm and the Trace-DF algorithm have improved in terms of accuracy, false alarm rate, and time performance.

4.2 Time-Consuming Comparison of Experimental Simulation with Different Algorithms

How much time is spent is a key indicator to measure the practicability of an algorithm. By comparing the running time of the three algorithms in the problem of Internet information collection and data analysis, the algorithm with less time consuming is more practical. Table 4 and Fig. 6 are obtained by calculating the consuming time using the three algorithms.

Figure 5 compares the running time of the three algorithms on the problem of Internet information collection and data analysis. The less time-consuming algorithm is more practical. In summary, the optimised model performance is more in line with the inspection requirements, the resolution of the experiment is optimised, the action data obtained by the detection can be stored more accurately, and the detection time is consumed, so that the model can better meet the timeliness of detection, and the reduction of the number of recognition errors is an essential improvement in the performance of the model. It shows that the Trace-DF algorithm is more suitable for information collection and data analysis.

Table 5
Data Comparison of Each Algorithm

Algorithm	Iterations	Experimental Time (min)	AA (%)
Hopfield	2000	15	70.5
RNN	2000	14	76.8
SOM	2200	12	66.4
VGG	2500	18	92.3
ELECTRED	2500	11	92.8
Trace-DF	3000	1 0	96.4

Table 6
Comparison of the Influence of Irrelevant Data on the Stability of the Algorithm

Algorithm	Proportion of Irrelevant Data (%)	Computation Time (ms)	Coefficient Size
Hopfield	50	584	4.4322
RNN	50	532	3.2955
SOM	50	548	3.2143
VGG	50	536	2.5857
Electred	50	536	2.0361
Trace-DF	50	536	1.8617

4.3 Data Comparison of Experimental Simulation with Different Model Algorithms

The Trace-DF network is used to compare the data set with traditional artificial intelligence algorithms, such as Hopfield, RNN, and VGG. The test data are shown in Table 5 and Fig. 7.

From these data, we can see that the convolutional neural network algorithm is better than other artificial intelligence learning algorithms in classification performance, and the convolutional neural network is more perfect than the traditional machine option algorithm in Internet information collection and data analysis. In order to better illustrate the advantages of the Trace-DF network, we will explain by discussing the stability of collecting and analysing data on the Internet. Variable parameters are involved in several artificial intelligence algorithms, and 50% of irrelevant parameters are selected here. Experiments are carried out on the amount of data, and the influence of variation interference on the prediction stability is analysed. The experiments are shown in Table 6, and the experimental results are shown in Fig. 8.

Note: The greater the correlation coefficient, the greater the interference intensity of the influence condition on the experimental variable.

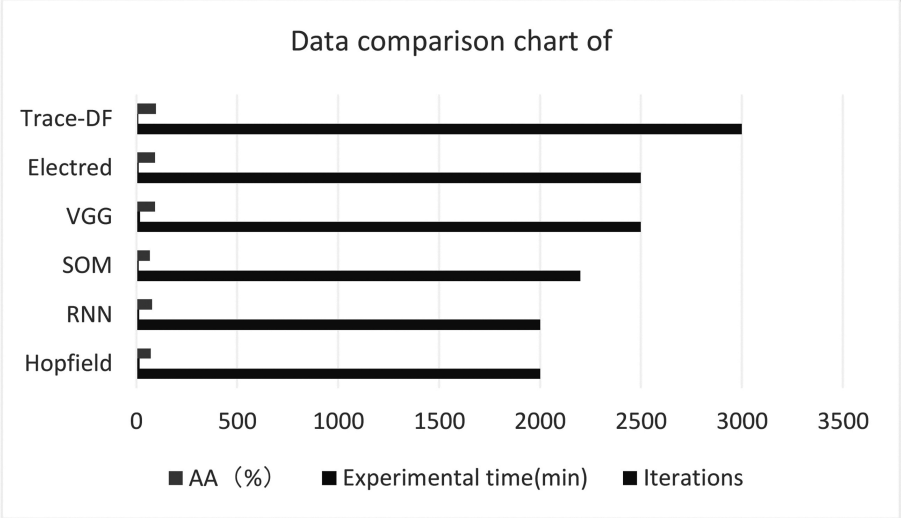


Figure 7. Data comparison of various algorithms in different experiments.

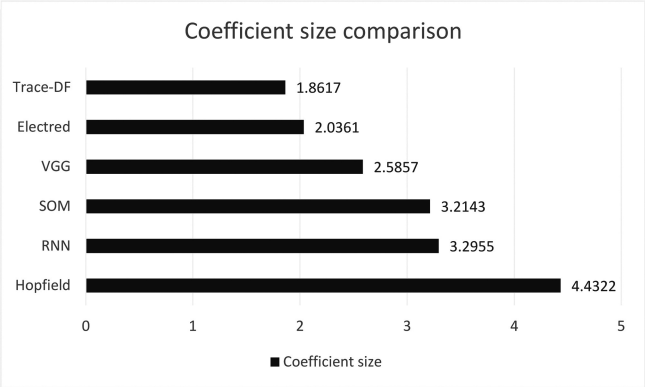


Figure 8. Data comparison diagram after interference experiment.

From the presented experimental results, it seems that the improved Trace-DF algorithm has the least impact under the interference of irrelevant data. Substantial improvements have been made in information collection and data analysis.

5. Conclusion

This paper analyses the current situation of Internet information collection and data analysis under big data. With the increasing amount of collected data and increasing complexity, more and more algorithm models are required. The traditional multi-dimensional network model for measuring is a relatively common algorithm, but for the analysis and processing of massive data, the speed of collecting information and data will be slowed down, and it may even affect the accuracy of the results. Therefore, the improved Trace-DF algorithm is used for calculation. The improved Trace-DF algorithm will be more stable and accurate for information collection and data analysis, and there will be no large error in the result rate. The future path of algorithm development should pay more attention to this point and improve the efficiency of the algorithm when analysing and processing data at the same time; consider the multi-thread processing method, strengthen the computing power of Internet information collection and data analysis, improve work efficiency, and ensure the user's network. The user experience reduces the workload of users to view information and solves the inefficiency of information collection and data analysis by artificial intelligence in the past. Compared with the traditional manual collection of information and analysis data and the analysis of a small amount of data, the main advantage is reflected in the collection of multi-dimensional models of Internet measurement and the rigorous analysis of the information data required by the collected information data and the information data analysis technology algorithm after the collection. In our simulation experiment test, we mainly studied the test performance of the reasonable application of artificial intelligence technology algorithm to Internet information collection and data analysis. And when we sort out the complex and mostly useless data into simple and clear required data under artificial intelligence technology, we cleverly use the FM algorithm to eliminate the useless data, ensuring the data accuracy after information collection and data analysis, so that the evaluation results are more scientific and rigorous, close to the real situation.

It can be applied to the public transport industry, high-end technology manufacturing, and precision machinery manufacturing industry.

Metric algorithm analyses and processes massive data, which will slow down the speed of collecting information data, and may even affect the accuracy of the results, while the ELECTRED algorithm classifies and detects the relevant collected data, clarifies the level relationship of the collected data, and then analyses and sorts the target data. After the data is analysed and sorted, the related calculation amount can be reduced. The improved Trace-DF algorithm will be more stable and accurate for information collection and data analysis, and there will be no large error in the result rate.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Funding Statement

1. This work was sponsored in part by the Henan Province Higher Education Training Program for Young Core Teachers, under Grant No. (2021GGJS192);
2. This work was sponsored in part by 2022 school-level scientific research project of Shangqiu Institute of Technology (2022KYXM44).

References

[1] Z. Ghahramani, Probabilistic machine learning and artificial intelligence, *Nature*, *521*(7553), 2015, 452–459.

[2] J. Zou and H. Zhang, New key point detection technology under real-time eye tracking, *Mechatronic Systems and Control*, *47*(2), 2019 71–76.

[3] P.I. Frazier, W.B. Powell, and S. Dayanik, A knowledge-gradient policy for sequential information collection, *SIAM Journal on Control & Optimization*, *47*(5), 2007, 2410–2439.

[4] V.R. Prybutok, An introduction to statistical methods and data analysis, *Technometrics*, *31*(3), 2001, 389–390.

[5] J.M. Chambers, W.S. Cleveland, and B. Kleiner, Graphical methods for data analysis, *Biometrics*, *40*(2), 2017, 493–499.

[6] Z. Chen, O. Lam, A. Jacobson, and M. Milford, Convolutional neural network-based place recognition, 2014, *arXiv:1411.1509*.

[7] L. Churn-Jung, Belief, information acquisition, and trust in multi-agent systems—A modal logic formulation, *Artificial Intelligence*, *149*(1), 2003, 31–60.

[8] L. Guo and G. Iyer, Information acquisition and sharing in a vertical relationship, *Marketing Science*, *29*(3), 2010, 483–506.

[9] C. Ulu and J.E. Smith, Uncertainty, information acquisition, and technology adoption, *Operations Research*, *57*(3), 2009, 740–752.

[10] S. Guha, K. Munagala, and S. Sarkar, Information acquisition and exploitation in multichannel wireless networks, 2008, arXiv:0804.1724, Doi:10.48550/arXiv.0804.1724.

[11] F.D. Missier, D. Ferrante, and E. Costantini, Focusing effects in predecisional information acquisition, *Acta Psychologica*, *125*(2), 2007, 155–174.

[12] R. Chahrour, Public communication and information acquisition, *American Economic Journal: Macroeconomics*, *6*(3), 2014, 73–101.

[13] K. Dehnad, Density estimation for statistics and data analysis, *Technometrics*, 29(4), 2012, 495–495.

[14] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin, Bayesian data analysis, *Wiley Interdisciplinary Reviews Cognitive Science*, *1*(5), 2014, 658–676.

[15] R.W. Argyle, P.S., Bunclark, and J.R. Lewis, *Astronomical Data Analysis Software and Systems*. (San Francisco, CA: Astronomical Society of the Pacific, 2008).

[16] Z. Fan, L. Linna, and Y. Bingru, Design and Implementation of Web-based intelligent information collection and processing system. *Computer Engineering*, *33*(18), 2007, 3.

[17] L. Guibing and L. Hong, Research on intelligent data analysis technology under big data, *Science and Technology Information*, (30), 2013, 2.

[18] X. Lei, Exploring intelligent data analysis technology under big data, *Science and Technology Innovation Herald*, *11*(10), 2014, 1.

[19] D. Mingchun and L. Gang, Research on intelligent data analysis method based on artificial neural network, *Information Technology and Informatization*, *15*(5), 2008, 3.

[20] M. Qingchao and L. Yongliang, Research on intelligent data analysis system, *Modern Commerce and Industry*, *20*(11), 2008, 2.

[21] Z. Xuewei and J. Xiuli, Research on data analysis under neural network, *RTVU*, *14*(4), 2009, 3.

[22] D. Yihua and H. Junchuan, Design and preliminary realization of general internet information collection system, *Computer Application Research*, *12*(1), 2005, 192–194+215.

[23] X. Xin, H. Zhongqing, and D. Sanhong, Design and implementation of internet overseas Chinese information collection system, *Modern Library and Information Technology*, *26*(7-8), 2010, 95–101.

[24] X. Dawei and W. Yanbo, Research on Internet user behavior data collection and analysis, *Modern Information Technology*, *3*(6), 2019, 14–16.

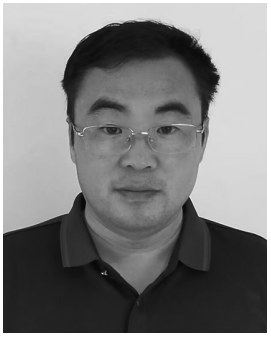
[25] W. Hui and P. Yunhe, From knowledge representation to representation: Progress in artificial intelligence epistemology, *Computer Research and Development*, *37*(7), 2000, 7.

[26] L. Wang, Automatic tracking algorithms based on wearable technology, *Mechatronic Systems and Control*, *201*(1), 2022, 50.

[27] Y. Li, R. Zhou, and F. Liang, Application research of cloud computing in electricity marketing field measurement remote acquisition system. *Mechatronic Systems and Control*, *50*(1), 2022, 55–60.

[28] W. Wu and Y. Wei, Guiding unmanned aerial vehicle path planning design based on improved ant colony algorithm, *Mechatronic Systems and Controls*, *49*(1), 2021, 48–54.

Biographies



Yongtao Zhang was born in 1982 in Henan, China. He received the bachelor's degree from Changjiang University in 2005. From 2005 to 2011, he worked with Shangqiu Institute of Technology. He received the master's degree from Henan University of Technology in 2014. Currently, he works with Shangqiu Institute of Technology. He has published ten papers, two of which have been included in SCI. His research interests include computer applications and the Internet of Things.

Na Yang was born in 1984 in Henan, China. She received the bachelor's degree from Luoyang Normal University in 2006, and the master's degree from Southwest Jiaotong University in 2009. Currently, she works with Shangqiu Institute of Technology. She has published nine papers, two of which have been included in SCI. Her research interests include digital signal processing and computer applications.