
Credibility Evaluation of Web Big Data Information Based on Particle Swarm Optimization

Nannan Zhao

*School of Computer Science and Engineering, Guangdong Ocean University at
Yangjiang, China
E-mail: znn@gdou.edu.cn*

Received 24 September 2021; Accepted 19 October 2021;
Publication 07 January 2022

Abstract

In order to improve the credibility evaluation effectiveness of web big data information, the improved particle swarm optimization is established. Firstly, framework of web big data is designed to include web big data source, data storage, data processing and data analysis. The global credibility calculation formula of whole web is established. Secondly, the improved particle swarm algorithm is constructed through updating weight and training factor, introducing cross and mutation operations into the algorithm, and improving population diversification based on mountain climbing algorithm. Finally, ten sample industries and one hundred sample stocks are selected to carry out experiment analysis, and the results show that the proposed algorithm can effectively distinguish trusted and untrustworthy records. And the proposed algorithm of calculating credibility of web big data has high reasonableness.

Keywords: Credibility evaluation, web big data, improved particle swarm algorithm.

1 Introduction

With rapid development of the web and communication technology, the information of Web has been growing explosively, and the Internet has stored a great number of information. The domain oriented deep Web big data source, which concludes rich structured data, has become an important means to obtain information. On the one side, all kinds of Web social media have become a main platform of transmitting information. On the other side, the challenges of Web big data brought by the development of Internet perplexes the effective utilization of big data. Big data has become the basic strategic resources in China, as it can promote the economic transformation and development. The government can apply big data to improve the urban competitive advantage and enhance governance capability. All kinds of industries and fields have promoted big data project. Web application safety is an important protection object of various industries, and currently more and more business has migrated to the Internet for processing due to the development of Internet, and there are more attacks on application system. The known defense way cannot completely avoid the malicious attacks in the world, including fraud, continuous penetration attack and so on. Taking the stock market as an example, more and more investors start to use the Internet to get investment information and find ideal investment channels, but facing complex big data on Web and thousands of released safety big data, investors cannot correctly identify the credibility of all safety information.

2 Literature Review

In recent years, the credibility of Web big data has been concerned by some scientists, Hongjian Guo proposed the tool of Web information data mining based on topic crawler [1]. Kapil Madan and Rajesh Bhatia presented a novel method to explore the unidentified deep web pages [2]. The credibility of Web big data involves the following problems:

The first problem is the privacy disclosure of user. Generally, investors need to offer an amount of personal information in the investment process. Therefore, some criminals collect information of investors by publishing false safety information, so as to get the personal information of investors and transfer the information to the third party, and the information of investors is disclosed. What is more, criminals used the personal information collected to fill in the credit card application form, and then swipe the card or withdrawn cash, and the investors has huge losses. Therefore, it is necessary to calculate the credibility of web big data.

The second problem is the optimal selection of Web big records. In the investment of the securities, the number of securities is far beyond the ability range of grasping the securities. Faced with an unknown security, investors often need to spend a lot of time to query whether the security is a security consistent with its description. Only when users understand the credibility of this security information, they can make correct investment according to actual situation in other fields, such as online shopping, booking air tickets, transferring second goods, and the credibility of web big data record is also a critical problem on users' own interests. Therefore, the credibility of web big data records widely exist in many fields of Web.

In order to solve the credibility of web big data record, the big data records in securities are analyzed, and the characteristics of web big data are obtained that are listed as follows [3]:

The higher the credibility of the Website published by a security information is, the higher the credibility of the security information is. The security information published on the Website of state-owned bank with high credibility has high credibility. The more reprinting times a security information is, the higher the credibility of this security information is.

If these characteristics of web big data are ignored, the credibility of a web big data record cannot be judged. In order to improve the effectiveness of evaluating web big data, the intelligent algorithm is an effective tool to evaluate the credibility of web big data, genetic algorithm and particle swarm algorithm are critical intelligent algorithms to carry out evaluate credibility of Web big data, the genetic algorithm has disadvantages of low convergence speed, and the particle swarm algorithm is easy to fall into local optimum, and has premature problem. Therefore, the particle swarm algorithm is improved to enhance the evaluation precision and efficiency [4].

Framework of Web Big Data

In order to ensure the credibility of Web big data, a whole platform is designed that integrates storage, processing, searching, visualization and data mining functions. The proposed platform can process and store various data, and integrate data collected, it can confirm the credibility of data, and can achieve data modeling and visualization operations [5].

The first step of constructing Web big data platform is to collect proper data, and the big data sources include data produced by machine, data generated by people, data in relational databases. The Web application system logs include Apache, NGINX and so on, the Web traffic packets and alarm

information of Web application firewall system should be collected. All kinds of data should be stored in database, the constructed database can support GB, TB and even ZB data sets. Therefore, in data storage, the distributed file processing system should be used as framework of Web big data storage, the pressure of writing can be alleviated based on sub database and sub table. The credibility analysis of Web big data can be carried out based on intelligent algorithm. The analysis results have visualized and displayed externally [6].

(1) Web big data source

The effective date of supporting credibility analysis should be extracted, which mainly includes outer and inner data, the outer data has particularity of security, and the security vulnerabilities mainly come from the information released by department. The outer date can disclose the credibility of system, and it can be associated with inner data to evaluate the credibility of system.

(2) Data storage

The relational database requires the united data format. When Web big data has no enough storage performance, only the method of updating CPU, memory and hard risk can be used to make up this disadvantage, but these methods can generate the performance bottlenecks. The best solution is to connect computing clusters through adding nodes, The Hadoop distributed file system (HDFS) in open source technology is used as the Web big data storage framework, the real-time query speed can be obtained on the large scale servers. It is composed of a named node and multiple data nodes, where name node saves attributes of all files, record data node in which each file block is stored. Data node storage file block is a cluster that includes a name node and a large number of data nodes to achieve the distributed storage of web big data [7].

(3) Data processing

Web big data has five different processing technologies: the first technology is to execute the batch processing of Hadoop MapReduce. The second technology is to access and store the big data from real time NoSQL database. The third technology is to apply the mature large scale parallel processing system optimized by business to process web big data. The fourth technology is compute and process big data in memory such as spark. The fifth technology is real time processing method of data flow, which processes big data in Storm.

MapReduce belongs to an open distributed framework, and it includes Map and Reduce, Map divides the initial input into multiple tasks, and inputs the results to reduce and synthesize the final results to output. Map and Reduce tasks are distributed on cluster nodes, and the results are stored on HDFS. MapReduce framework and HDFS run on the same set of nodes. This configuration enables the stored data to execute tasks efficiently on the node, so that the whole Web cluster can be used efficiently [8].

Spark is the alternative processing system, which is distributed memory computing system. It refers to the data processing on memory, and the cost of memory decreases due to development of memory technology. The data processing speed can be improved based on computing site using a large number of memory. The computing speed of memory is quicker than visiting speed of disk, and it can better obtain, summarize and analyze web big data. Spark is also an open source project of Apache foundation, and it divides every job into a series of tasks and sends them to a cluster composed of several servers. It submits the web big data to memory for calculation. In this way, Spark needs to write to the disk once, instead of constantly writing files on the disk, so it achieves better processing capability [9].

(4) Web big data analysis

In Web, there are a lot of data sources in every field, and information exists as record in data source. Generally, different data sources in the same field are isolated from each other, and there is no direct connection. But, the same record may be appearing repeatedly in the multiple different data sources, therefore content overlap may exist between data sources. Moreover, although there is no direct connection between records in different data sources, indirect connection can also be established for records through the contact relationship between websites. According to the overlap between data sources and the relationship between records and websites, this paper proposes a method to build a new credibility Web for a record.

The credibility Web includes two kinds of vertices and three types of edges for a record. The two kinds of vertices of the credibility Web are listed as follows: The first vertices is site vertices (v^s): it is the website concluding data record. For different fields, this kind of vertices refers to a different website that releases information of different fields. The second vertices is record vertices (v^r), it is the data record on website, all record vertices in the proposed credibility Web represent the same content of data, but they may be scattered in different data sources [10].

The credibility Web includes three edges. The first edge is the internal link edge (e^i). If a record belongs to a data source, in credibility Web the record vertex corresponding to this record and the site vertex corresponding to this data source construct the direct edge. The direction of the edge is from the site vertex to the record vertex. This directed edge represents the relationship between the website and records of it, so it is called internal link edge [11].

The second edge is the outer link edge (e^o): it can represent link relationship between records, or between record and outer data source. If a record contains a link to the external data source, or a record contains a link to a record of other database, then in Web there is directed edge between its corresponding vertices, and the direction of edge is the same as that of link.

The third edge is identification edge (e^r): if two records belonging to different data sources are verified by entity recognition technology, and they are considered to represent the same entity, and then there is an undirected edge between two record vertices [12].

The novel credibility Web established based on link relationship and entity identification technology is constructed, which is shown in Figure 1. The initial credibility value should be given to every vertex of Web before the credibility value of a record is calculated.

The credibility of a vertex can be transferred to its adjacent vertices through directed and undirected edges. This propagation behavior is carried out among all nodes of the whole web. The global credibility of whole Web is calculated, which is the credibility value of the record in Web.

The propagation rate of credibility of different types of edge is also different. The propagation rate of inner link edge is defined by $\phi(e^i)$, the propagation rate of outer link edge is defined by $\varphi(e^o)$, and the propagation rate of entity identification edge is defined by $\psi(e^r)$. The vertex credibility after n times credibility value propagation can be used to calculate the vertex credibility value after $n + 1$ times propagation, and the corresponding equation is listed as follows [13]:

$$C_v^{n+1} = \lambda BC_v^n + \frac{1 - \lambda}{|R|} s \quad (1)$$

where C_v^n represents the credibility value vector of vertex after n times credibility propagation, B is a $m \times m$ matrix, an element in B is defined by $B_{ij} = \phi(e^i)$, d represents the damping coefficient, R represents any sub set of all vertices, $r = (r_1, r_2, \dots, r_m)^T$ is a basic vector, if $v_i \in R$, $r_i = 1$.

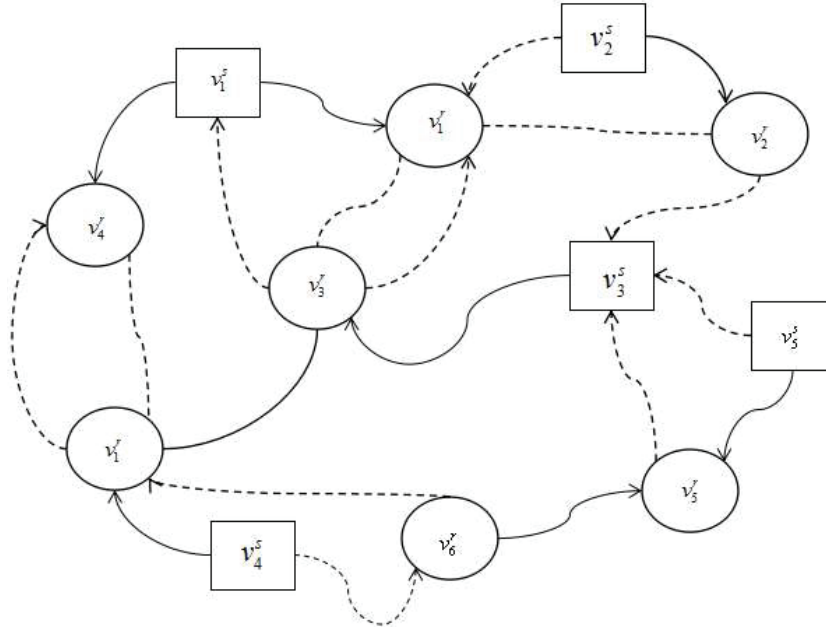


Figure 1 Structural diagram of credibility web.

An arbitrarily small vector is set as δ , when the following conditions are met, the iteration stops [14].

$$|C_v^{m+1} - C_v^m| < \delta \tag{2}$$

The global credibility of whole web is calculated by

$$C_g = \sum \omega_n(v_i)r_v(v_i) \tag{3}$$

where $\omega_n(v_i)$ represents the weight of different vertices, and it is standardized by the following equation:

$$\omega_n(v_i) = \frac{\omega(v_i)}{\max\{\omega(v_k), k = 0, 1, 2, \dots, m\}} \tag{4}$$

The Equation (4) can effectively avoid malicious information publishers from repeatedly publishing and reprinting the same false recruitment information on multiple websites. If the malicious publisher wants to increase the number of records in this way to defraud a certain credibility value, it is impossible to obtain a high global credibility value. This method can calculate a credibility of web big data that is suitable for real situation.

Improved Particle Swarm Algorithm

In order to evaluate the credibility of web big data, the particle swarm algorithm is used in this research. The basic thinking of the particle swarm optimization algorithm is to randomly initialize a group of particles without mass and volume in feasible solution space, and each particle represents a feasible solution of the optimization problem. Each particle flies at a certain speed in the feasible solution space. The velocity decides moving direction and distance of the particle, which is adjusted according to flight experience of the particle itself and the flight experience of the entire group. The quality of particle is decided by the preset fitness function. Generally, the optimal solution is obtained through tracking the location of current optimal particle and searching and updating it for many times. In each generation, the particle can update its location through tracking two extreme values, one is the individual extreme solution that is defined by P_{best} , and the other is global extreme solution that is defined by G_{best} [15].

In D dimension feasible searching space, the searching population is made up of M particles, and they can fly in a certain speed in space, which are defined by

$$X = \{X_1, X_2, \dots, X_M\} \quad (5)$$

The flying speed of i th particle is defined by

$$V_i = \{v_{i1}, v_{i2}, \dots, v_{iM}\} \quad (6)$$

The location of i th particle is defined by

$$L_i = \{l_{i1}, l_{i2}, \dots, l_{iM}\} \quad (7)$$

The optimal location of individual is defined by

$$p_i = \{p_{i1}, p_{i2}, \dots, p_{iM}\} \quad (8)$$

The global optimal location is defined by

$$g = \{g_1, g_2, \dots, g_M\} \quad (9)$$

The original particle swarm algorithm is expressed by [16]

$$v_{id}^{k+1} = \omega v_{id}^k + \kappa_1 r_1 (p_{id}^k - l_{id}^k) + \kappa_2 r_2 (g_{id}^k - l_{id}^k) \quad (10)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (11)$$

where κ_1 and κ_2 are training factor, r_1 and r_2 are random number ranging from 0 to 1, v_{id}^{k+1} denotes the d th dimensional velocity component in the $k + 1$ th time iteration, v_{id}^k denotes the d th dimensional velocity component in the k th time iteration, l_{id}^{k+1} denotes the d th dimensional location component in the $k + 1$ th time iteration, l_{id}^k denotes the d th dimensional location component in the k th time iteration, p_{id}^k denotes the d th dimensional location component corresponding to optimal fitness value of i th particle in the k th time iteration, g_d^k denotes the d th dimensional location component for whole particle swarm in the k th time iteration, ω is the inertia weight [17].

The velocity of particle is made up of three parts: the first part is the velocity of particle in the last iteration, which shows the memory ability of particle. The particle carries out inertial motion according to previous velocity. The second part is cognitive portion, which is the thinking of particles about themselves and the comparison between the current optimal location and best location of particles [18, 19]. The third part is social portion, which is the comparison between current position and best position of whole population, and it illustrates information share and cooperation between particles. Therefore, the particle can jump out of the local optimal solution and find out global solution quickly according its experience.

In particle swarm algorithm, κ_1 can be used to control the distance of individual particles flying to the individual optimal location, κ_2 can be used to control the distance of individual particles flying to the global optimal location, the proper value of κ_1 and κ_2 can improve computing efficiency of particle swarm algorithm, and avoid to fall into local optimum. In general, the searching space has a certain range, therefore the location of particle should range from l_{\min} to l_{\max} , the location of particle is also limited by maximum velocity, the velocity of particle should be taken as v_{\max} when the velocity is larger than . If the maximum velocity of particle is too small, the global searching ability of particle will reduce, and if the maximum velocity of particle is too large, the particle can cross the optimal location, and the algorithm cannot reach the optimal solution [20].

According to the past practice, the particle swarm algorithm has been concerned by many scholars. It has been applied in many fields, because of its easy operation, simple conception, and few affecting factors. However, it has also some defects, for example, it is easy to fall into local optimum in the later stage of iteration, and is difficult to jump out of local extreme value; it has poor global searching capability, and the accuracy and efficiency

of obtaining global solution are poor. Therefore, some measurements are put forward by some scholars, and there are three measurements of avoiding the defects of traditional particle swarm algorithm.

(1) The parameters in algorithm can be improved. For example, the inertia factor is regulated, the training factor is improved, the population scale is optimized and the ending condition of algorithm is set. The inertia weight is an important parameter of particle swarm algorithm, proper selection of the inertia weight can be favorable for balancing global and local searching performance of algorithm. When the inertia weight is small, the global search ability is weakened and the local search ability is strengthened, and a more accurate solution can be obtained; When the value is large, the global search ability is strengthened, the local search ability is weakened, and the convergence speed is accelerated. Reasonable selection of inertia weight can improve the search performance and optimization ability of the algorithm and reduce the number of iterations of the algorithm. However, when the inertia weight is large, it is not conducive to the accurate solution of the algorithm, while when it is small, it will reduce the convergence speed of the algorithm and easy to fall into local optimization. The nonlinear dynamical strategy based on arctangent function of inertia factor is used in this research, and the corresponding expression is listed as follows:

$$\omega(t) = (\omega_s - \omega_e) \cdot \arctan \left(1.67 \times \left(1 - \left(\frac{T}{T_{\max}} \right)^\eta \right) \right) + \omega_e \quad (12)$$

where ω_s is the starting value of inertia factor, ω_e is the ending value of inertia factor; η is the controlling factor, which controls smooth degree of changing curves of inertia factor and period. T_{\max} is the maximum period. This improvement expression can ensure optimization stability of function fitness, the inertia weight decreases nonlinearly with the increase of the number of iterations, and the value decreases slowly in the early stage of the iteration and quickly in the later stage. This makes the algorithm ensure not only the convergence speed in the early stage but also the search ability in the later stage.

The training factors κ_1 and κ_2 are also important factors of algorithm. Ideally, we hope that particle swarm optimization can have global search ability in the initial stage of the algorithm and search in the whole optimization space without falling into local minimum; In the later stage of the algorithm, it can converge to the global minimum as soon as possible, speed up the convergence speed and improve the convergence accuracy. Therefore,

this paper achieves this goal by dynamically adjusting the size of κ_1 and κ_2 learning factors. The basic operation methods are as follows: in the initial stage of the algorithm, there is a large cognitive part and a small social part, which helps to improve the global search ability; In the later stage of the algorithm, there is a small cognitive part and a large social part, which is more helpful for the algorithm to converge to the global optimal solution and improve the accuracy and convergence speed of the algorithm. The calculating formulas of κ_1 and κ_2 are listed as follows:

$$\kappa_1 = (\kappa_s - \kappa_e) \cdot \frac{I_{iter,max} - I_{cur}}{I_{iter,max}} + \kappa_e \quad (13)$$

$$\kappa_2 = 4.5 - \kappa_1 \quad (14)$$

Where $I_{iter,max}$ denotes the maximum iteration times, I_{cur} denotes the current iteration times, κ_s denotes the starting value of training factor, κ_e denotes the ending value of training factor.

(2) The particle swarm algorithm can be combined with other intelligent algorithm to learn from strong advantages of other algorithm and close the gap. The crossover operation is used to illustrate the characteristics of information exchange between particles, and generation of crossover particles is achieved by applying roulette method, and the particles in the next generation is expressed by

$$X_i^{t+1} = \rho_i X_i^t + 1 - \rho_i X_j^t \quad (15)$$

$$X_j^{t+1} = (1 - \rho_i) X_i^t + \rho_i X_j^t \quad (16)$$

Where X_i^t and X_j^t are two cross particles in t th time iteration. X_i^{t+1} and X_j^{t+1} are particles in the next generation generated after crossing. ρ_i is the random number ranging from 0 to 1, and when it is close to 0 or 1, the population contraction velocity is slow; and when it is close to 0.5, the population contraction velocity is quick.

The single point mutation is applied in this research, the k th dimensional mutation point in mutation population is selected based on random number method, and the initial speed of particle is generated, which is expressed by

$$v_d^k = v_{\min} + rand \cdot (v_{\max} - v_{\min}) \quad (17)$$

where $rand$ represents the random number, $rand \in [0, 1]$.

(3) Population diversification improvement

Particle swarm optimization algorithm is a simple, efficient and practical intelligent optimization algorithm, and the advantage of it is quick convergence speed. However it is easy to fall into local optimum. The mountain climbing operation is introduced into the particle swarm algorithm, and the diversity of population is increased, and the local search capacity of algorithm is enhanced. Mountain climbing algorithm is an optimal technology by applying neighborhood search method to search in the direction that can enhance the quality of solutions, therefore the local search capacity has been improved.

The improved algorithm carries out mountain climbing operation for the new particle generated in each iteration of algorithm, because the mountain climbing algorithm has characteristics of strong local search ability, and the fitness degree of particle is improved. According to the features of real number coding, the neighborhood of the current solution is constructed by one-dimensional mutation. For the current particle, the dimension number is randomly selected and the value of the selected dimension in the position vector is randomly set, that is, the position vector of the particle is mutated in one dimension to obtain a new position vector. Then the fitness is calculated. If the new position vector is better than the original position vector, accept the new position vector, otherwise reject the new position vector. By increasing the mountain climbing operation, the diversity of particles can be effectively increased and the local search ability can be enhanced.

The one dimensional mutation method applied in this research is better than transposition method, reversal method and inserting method. Mountain climbing method is an algorithm depending on neighborhood search. In mountain climbing algorithm, new solutions are constructed by selecting points in the neighborhood. The method of determining neighborhood and selecting points is an important part of mountain climbing algorithm. Neighborhood point selection methods are commonly used to solve combination optimization problems including transposition method, reversal method and insertion method. Compared with transposition method, reversal method and insertion method, the one-dimensional mutation method used in this paper increases the possibility of adjusting the number of vehicles and expands the range of neighborhood search.

The fitness function of algorithm is used to compute the utility of web service composition, which is defined by *Fit*, the fitness value of different service composition is used to evaluate the good and bad of service

composition, and the fitness function is expressed by

$$Fit = \sum_{r=1}^M \omega_r \cdot J_{cs}^r \quad (18)$$

where ω_r denotes the weight of r th index, M denotes the total number of indexes, J_{cs}^r denotes the r th index.

Case Study

In order to verify the effectiveness of the proposed algorithm, simulation analysis is carried out. Ten sample industries and one hundred sample stocks are selected. To reflect the mutual influence, the selected ten sample industries have upstream and downstream relationships. Ten selected sample industries are more important and concerned industries in the near future, with large market fluctuations and frequent news. In this research, data acquisition, analysis and prediction are realized based on the recent time, and the real verification is executed based on results and subsequent situations. Through collecting all relevant information events in the half month from May 1, 2020 to May 15, 2020, there are 11387 relevant information events after duplication operation, with an average of about 759 per day. The daily data distribution is listed in Table 1.

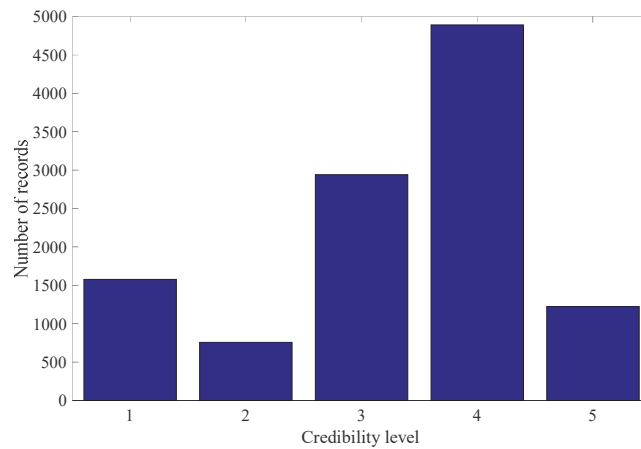
Before simulation analysis, a credibility network is automatically constructed for each securities record. The initial credibility value of site vertex is set to the PR value of the site, and the initial credibility value of record vertex is equal to zero. If a website is an open platform that allows free release of information, its initial credibility is set as the product of PR value and a discount coefficient. The discount coefficient is got by statistical analysis, and it ranges from 0 to 1.

The credibility value of web big data is divided into five intervals, including $[0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$, $[0.8, 1.0]$, these five intervals represent five credibility level. The higher the level is, the more credible is. The number of records includes in the credibility levels 1–5 as shown in Figure 2, which is 1577, 757, 2940, 4890, and 1223 respectively.

As seen from Figure 2, all untrusted records are identified. Moreover, the number of second level records is relatively small, which is the through of the whole trend, and the results show that the proposed algorithm can effectively distinguish trusted and untrusted records. The distribution of the credibility of recruitment records at levels 2–5 shows a trend of normal distribution, the

Table 1 Daily data distribution of all relevant events

Date	Market	Industry	Individual Share
May 1	35	42	1032
May 2	13	89	754
May 3	35	99	945
May 4	43	53	752
May 5	53	62	834
May 6	44	56	845
May 7	55	76	730
May 8	22	68	894
May 9	46	68	553
May 10	42	75	442
May 11	37	42	653
May 12	30	36	885
May 13	41	104	765
May 14	42	59	760
May 15	39	70	543

**Figure 2** Number of records with different credibility level.

results are consistent with the real situation of Web big data, and the results can also give a reasonable and appropriate evaluation value for credibility of records.

To verify the accuracy of credibility evaluation results, fifteen investors are invited to carry out experiment, the system randomly assigns 150 securities records to each user, and the credibility value of Web big data based proposed algorithm. The investor evaluates whether the credibility value

Table 2 Evaluation results of investors for credibility of web big data

Investor	Number of Records			
	Reasonable	High	Low	Unreasonable
1	143	5	1	1
2	142	6	2	0
3	140	5	2	3
4	146	1	2	1
5	145	2	3	0
6	140	4	3	3
7	144	3	2	1
8	145	2	2	1
9	146	2	2	0
10	144	3	1	2
11	148	1	1	0
12	149	0	1	0
13	147	2	1	0
14	146	2	1	1
15	148	1	1	0

Table 3 Evaluation accuracy of different models

Credibility Level	Traditional Model	Proposed Model in This Research
1	88.6%	95.3%
2	89.4%	96.3%
3	87.9%	97.5%
4	88.2%	97.4%
5	90.1%	98.4%

of every record is reasonable according to judgment, the evaluation levels include reasonable, high, low and unreasonable, and the evaluation results are shown in Table 2.

As seen from Table 2, the investors think that the proposed algorithm of calculating credibility of web big data has high reasonableness. The reasonable number of credibility evaluation record for every investor is overpass 140, and the high, low and unreasonable number of credibility evaluation records is less than 5. Results show that the proposed method can provide valuable reference information for investors based on calculated credibility of web big data.

In order to verify the effectiveness of proposed model, the traditional evaluation is also used to carry out evaluation analysis, and the compared results are listed in Table 3. As seen in Table 3, the proposed model in this research has higher evaluation accuracy than traditional model.

Conclusions

The improved particle swarm algorithm is established to evaluate credibility of web big data, and the credibility network is constructed for every record of web. Results show that the proposed method can effectively evaluate the credibility of web big data, which can identify the false information of web big data, and can provide the valuable information for users, and the web big data credibility evaluation platform is designed to have a wide application range.

References

- [1] Mohamad FirdausA Aziz, Salama A Mostaf, Cik Feresa Mohd. Foozy, Mazin Abed Mohammed, Mohamed Elhoseny, Abedallah Zaid Abualkishik, Integrating Elman recurrent neural network with particle swarm optimization algorithms for an improved hybrid training of multidisciplinary datasets, *Expert Systems with Applications, Expert Systems with Applications*, 2021, 183(11):115441.
- [2] Ozgur Kisi, Payam Khosravini, Salim Heddami, Bakhtiar Karimi, Nazir Karimi, Modeling wetting front redistribution of drip irrigation systems using a new machine learning method: Adaptive neuro-fuzzy system improved by hybrid particle swarm optimization-Gravity search algorithm, *Agricultural Water Management*, 2021, 256(10):107067.
- [3] Payam Parsa, Hosein Naderpourb, Shear strength estimation of reinforced concrete walls using support vector regression improved by Teaching-learning-based optimization, Particle Swarm optimization, and Harris Hawks Optimization algorithms, *Journal of Building Engineering*, 2021, 44(12):102593.
- [4] Wang Rong and Feng Yue, Evaluation research on green degree of equipment manufacturing industry based on improved particle swarm optimization algorithm, *Chaos, Solitons & Fractals*, 2020, 131(2):109502.
- [5] Chao Wang, Jin Ming Koh, Tiantang Yu, Neng Gang Xie, Kang Hao Cheong, Material and shape optimization of bi-directional functionally graded plates by GIGA and an improved multi-objective particle swarm optimization algorithm, *Computer Methods in Applied Mechanics and Engineering*, 2020, 366(7):113017.
- [6] Haojie Ding, Xingsheng Gu, Improved particle swarm optimization algorithm based novel encoding and decoding schemes for flexible job

- shop scheduling problem, *Computers & Operations Research*, 2020, 121(9):104951.
- [7] Wenyang Zhang, A Probability Distribution and Location-aware ResNet Approach for QoS Prediction, *Journal of Web Engineering*, 2021, 20(4):1189–1228.
- [8] Zhuang Wang, Jiejun Cai, The path-planning in radioactive environment of nuclear facilities using an improved particle swarm optimization algorithm, *Nuclear Engineering and Design*, 2018, 326(1):79–86
- [9] Jianhu Gong, Design and Analysis of Low Delay Deterministic Network Based on Data Mining Association Analysis, *Journal of Web Engineering*, 2021, 20(2):513–532.
- [10] Zhiyuan Ouyang, Yu Liu, Sheng-Jia Ruan, Tao Jiang, An improved particle swarm optimization algorithm for reliability-redundancy allocation problem with mixed redundancy strategy and heterogeneous components, *Reliability Engineering & System Safety*, 2019, 181(1):62–74.
- [11] Keyan Ma, Mingsheng Liu, Jili Zhang, An improved particle swarm optimization algorithm for the optimization and group control of water-side free cooling using cooling towers, *Building and Environment*, 2020, 182(9):107167.
- [12] Han Zhengtong, Gu Zhengqi, Ma Xiaokui, Chen Wanglin, Multimaterial layout optimization of truss structures via an improved particle swarm optimization algorithm, *Computers & Structures*, 2019, 222(10):10–24.
- [13] Guan Guan, Yang Qu, Gu Wenwen, Jiang Wenying, Lin Yan, Ship inner shell optimization based on the improved particle swarm optimization algorithm, *Advances in Engineering Software*, 2018, 123(9):104–116.
- [14] Asgarali Bouyer, Abdolreza Hatamlou, An efficient hybrid clustering method based on improved cuckoo optimization and modified particle swarm optimization algorithms, *Applied Soft Computing*, 2018, 67(6):172–182.
- [15] Lingling Chen, Qi Li, Xiaohui Zhao, Zhiyi Fang, Furong Peng, Jiaqi Wang, Multi-population coevolutionary dynamic multi-objective particle swarm optimization algorithm for power control based on improved crowding distance archive management in CRNs, *Computer Communications*, 2019, 145(9):146–160.
- [16] Zhen Wang, Jihui Zhang, Shengxiang Yang, An improved particle swarm optimization algorithm for dynamic job shop scheduling problems with random job arrivals, *Swarm and Evolutionary Computation*, 2019, 51(12):100594.

- [17] Ahmad Muhaimin Ismaila, Mohd Saberi Mohamad, Hairudin Abdul Majid, Khairul Hamimah Abas, Safaai Deris, Nazar Zaki, Siti Zaiton Mohd Hashim, Zuwairie Ibrahim, Muhammad Akmal Remli, An improved hybrid of particle swarm optimization and the gravitational search algorithm to produce a kinetic parameter estimation of aspartate biochemical pathways, *Biosystems*, 2017, 162(12):81–89.
- [18] Hathiram Nenavath, Dr Ravi Kumar Jatoth, Dr Swagatam Das, A synergy of the sine-cosine algorithm and particle swarm optimizer for improved global optimization and object tracking, *Swarm and Evolutionary Computation*, 2018, 43(12):1–30.
- [19] Mădălina Eraşcu, Flavia Micota, Daniela Zaharie, Scalable optimal deployment in the cloud of component-based applications using optimization modulo theory, mathematical programming and symmetry breaking, *Journal of Logical and Algebraic Methods in Programming*, 2021, 121(6):100664.
- [20] Igor Granado, Leticia Hernando, Ibon Galparsorom Gorka Gabiña, Carlos Groba, Raul Pallezo, Jose A. Fernandes, Towards a framework for fishing route optimization decision support systems: Review of the state-of-the-art and challenges, *Journal of Cleaner Production*, 2021, 320(10):128661

Biography



Nannan Zhao received her B.Sc. and M.E. degrees in Computer Application Technology from Liaoning Technical University, China. She was awarded the title of Outstanding Teacher of Private Education in Guangdong Province. She won the Award of Guangdong Excellent Online Teaching Case, the special prize of university-level teaching Achievement Award, the Excellence Award of the 4th Guangdong Provincial College (undergraduate) Young Teachers' Teaching Competition, the third prize of micro class of higher Education Group in Guangdong Provincial Computer Education Software Evaluation Activity, and the third prize of scientific research project of Guangdong Provincial Finance Department. She has been an excellent member of Jiusan Society in Zhanjiang City from 2019.

