Big Data Trust Evaluation Based on D-S Evidence Theory and PageRank Model

Yu-Ling Chen[1,2], Yu-Jun Liu[5], Wei-Fa Zheng[3,4], Jing-Yao Chen[4]

[1] State Key Laboratory of Public Big Data, GuiZhou University, China
[2] College of Computer Science&Technologe, GuiZhou University, China
[3] Network Information Center, Guangdong University of Finance and Economics, China
[4] College of Engineering, Shantou University, China
[5] Technical Center of Beijing Customs, China
ylchen3@gzu.edu.cn, liuyjbj@163.com, wf@gdufe.edu.cn, 18jychen3@stu.edu.cn

Abstract

In view of the multi-dimensional attributes and uncertainties existing in the trust evaluation of big data nodes under the big data environment, this paper proposes a big data distributed collaborative trust management framework and a trust assessment model and a trust assessment model. On the basis of big data production environment, the new big data trust processing agent is creatively proposed to calculate and manage data source trust information and establish trust network through trust processing agent. In the calculation of direct trust, this paper adopts the multi-dimensional trust evaluation method of D-S evidence theory to evaluate the direct trust value of the trust agent to the big data source and uses dynamic weight to modify Dempster's rule of combination to avoid evidence conflict. When calculating the indirect trust value, this paper uses PageRank algorithm to calculate the recommended value of the trust processing agent. Finally, the integrated trust values of big data sources are obtained by combining direct and indirect trust. Experimental analysis shows that the trust model can make trust evaluation for big data sources accurately, has obvious ability to distinguish trust, and has strong robustness.

Keywords: Big data, Trust evaluation, D-S evidence theory, PageRank model

1 Introduction

In the era of big data, most data sources are provided to various big data systems in the form of database or online API. The generation process of big data system is complex, which requires parallel or series processing of multiple data to produce useful results. The 4V characteristics of big data make it hard to avoid problems such as inconsistency and lack of data in the process of producing, transmitting and receiving big data, that is, it is easy to produce unreliable data [1]. In addition, the stability and reliability of data sources is also an important guarantee for the production of big data. If there is no credible data environment, the use of big data will bring great risks and hidden dangers, misleading and harmful results to user decisions. Only the real and credible data environment is the basis for creating value. Therefore, the credibility calculation of big data sources should be carried out before the big data source to screen out the real and credible data or the data environment with high credibility and to eliminate the data with unreliability or low credibility.

There has been a lot of discussion about the definition of trust. Different scholars have given different interpretations and reached no consensus. Over the years, researchers have proposed many different expressions of the concept of trust from different perspectives. Social scientists tend to attribute trust to the decisive influence of external society, environment and organization on human behavior, and the influence of internal biological mechanism on individual decision-making and use game theory to study it [2]. For example, Abdul-Rahman et al. [3] defined "trust" as: trust refers to a person's ability to complete a specific task as scheduled, which reflects a personmoral level and social existence. Buskens [4] gave the following definition of "trust": trust mainly comes from previous interaction experience, and based on the interaction experience, select whether to trust the promise of the other party, and disclose personal privacy information to the other party. In the field of computer, the emergence and development of the concept of "trust" make up for the deficiency of traditional computer security technology and improve the security of network and software system. Wang and Vassileva [5] explained the concept of trust as a direct evaluation of an individual based on his own experience and a direct measurement of the credibility

*Corresponding Author: Wei-Fa Zheng; E-mail: wf@gdufe.edu.cn
DOI: 10.3966/160792642020072104006

and reliability of another individual.

At present, the research on trust mainly focuses on multi-agent system, P2P network, social network, cloud service and other fields. Trust measurement model based on fuzzy theory mainly contains [6], based on information entropy theory [7], based on the theory of evidence and probability statistics [8]. Josang [9] and Muil et al. [10] in their trust model using the Bayesian estimation theory, Josang in his model was proposed based on the Beta distribution function to describe the ideas of the a posteriori probability, the binomial events is given a based on positive and negative events for certainty probability density function, Muil et al. also used this method in his article, but they did not grade the trust, that is, did not distinguish the direct trust value from the recommended trust value. Wu et al. [11] proposed a multidimensional trust assessment method based on d-s evidence theory to evaluate the trust value of cloud service providers in view of the multidimensional attributes and uncertainties of trust assessment in the current cloud environment. Li et al. [12] proposed a trust evaluation model based on entropy weight method. In the fusion calculation process of multi-dimensional decision attributes, the information entropy theory was used to establish the classification weight of each decision attribute to avoid the weak adaptability of subjective judgment method in weight setting and ensure the effectiveness and objectivity of the recommendation trust evaluation decision.

However, there are not many researches on big data using trust theory. Johannes Sanger et al. is among the first to trust evaluation relates to one of the researchers in the field of big data, he big data in [13] the authenticity of the definition, judgment, evaluation and other issues put forward the research route, and from the objective or subjective, credible or deceive, reliable or unreliable the three dimensions of authenticity, and discussed the measurement and evaluation of the three dimensions of existing tools and techniques; [14] puts forward that from the perspective of big data security, one of the most important problems to be solved is the credibility of each source or information, and proposes a method to evaluate and quantify the trust level of information sources and information items by using a large number of literature citation rankings. This method uses the trust evaluation model for reference to the PageRank method of Google and realizes the literature rating as a social choice problem. Based on the traditional data trust analysis theory, Li et al. [15] constructed a hierarchical network model for dynamic big data trust analysis by adding weight parameters such as time factor and penalty factor. In this paper, a big data credibility computing framework is proposed. In this paper, a big data credibility computing framework is proposed. The direct trust is calculated by improving the D-S evidence theory model, the indirect trust is calculated by PageRank model, and the comprehensive trust is finally calculated.

2 D-S Evidence Theory

Evidence theory originated from the research work of A. P. Dempster, a mathematician at Harvard University in the 1960s, which used upper and lower limit probability to solve the problem of multi-valued mapping. His students G. Shafer theory of Evidence for further development, the introduction of trust function concept, formed a set based on the "Evidence" and "combination" deals with the problem of uncertainty reasoning according to the Mathematical method, and in 1976 published A Mathematical found of Evidence, it marks the Evidence Theory became a complete theory of dealing with uncertainty. In D-S evidence theory, the recognition framework Θ is used to represent the complete set of the object studied, and the elements in Q are mutually exclusive and discrete values. The set of all subsets based on the recognition framework Θ is called the power set of Θ, which is called 2^{Θ}. [16-18]

Definition 1: Basic Probability Assignment Function m (BPA). In the identification framework Θ, if m: $2^{\Theta} \to [0,1]$ exists, it is satisfied

$$\begin{cases} \mathrm{m}(\varnothing) = 0 \\ \sum\limits_{A \subseteq \Theta} m(A) = 1 \end{cases} \tag{1}$$

Then m is the basic probability distribution function of A, called mass function, so that A of m(A)>0 is called Focal elements.

Definition 2: Belief function. The trust function based on BPA m on the identification framework Q is defined as:

$$\mathrm{Bel(A)} = \sum_{\mathrm{B} \subseteq \mathrm{A}} \mathrm{m}(B) \tag{2}$$

Definition 3: likelihood function (Plausibility function), the likelihood function based on BPA m on the identification framework Q is defined as follows:

$$\mathrm{Pl(A)} = \sum_{\mathrm{B} \cap \mathrm{A} \neq \varnothing} \mathrm{m(B)} \tag{3}$$

Definition 4: reliability interval. The trust function Bel(A) and the likelihood function Pl(A) constitute the reliability interval [Bel(A), Pl(A)], which is used to express the degree of confirmation of the proposition. [0, 0] represents the negation of proposition A, [1, 1] represents the affirmation of proposition A. The description of the evidence reliability interval is shown in the Figure 1.

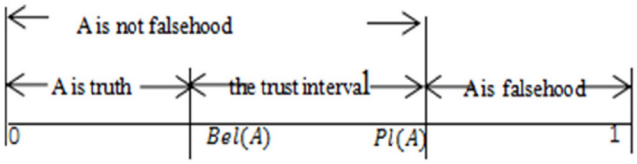


Figure 1. D-S reliability interval description

Definition 5: Dempster's combinational rule, also known as evidence combination formula. For two mass functions m_1 and m_2 on $\forall A \subseteq \Theta$ and Θ, Dempster's combination rule is:

$$m_1 \oplus m_2(A) = \frac{1}{K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \tag{4}$$

Where,K is the normalized constant

$$K = \sum_{B \cap C \neq \varnothing} m_1(B) \cdot m_2(C) \tag{5}$$

For $\forall A \subseteq \Theta$, identify the finite mass functions $m_1, m_2, \cdots, m_n$'s Dempster combination rule is:

$$\left(m_1 \oplus \cdots \oplus m_n\right)(A) = \frac{1}{K} \sum_{A_1 \cap \cdots \cap A_n = A} m_1\left(A_1\right) \cdots m_n\left(A_n\right) \tag{6}$$

where

$$K = \sum_{A_1 \cap \cdots \cap A_n \neq \varnothing} m_1\left(A_1\right) \cdot m_2\left(A_2\right) \cdots m_n\left(A_n\right) \tag{7}$$

K known as the normalization factor, reflect the degree of evidence conflict between, when K is close to zero, the smaller the degree of the conflict between body of evidence, the fusion results more accurate, when K is close to 1, the greater the degree of the conflict between body of evidence, the fusion results more accurate, when K is close to 1, the greater the degree of the conflict between body of evidence, the fusion result is not accurate, when K = 1, the contradiction between the evidence body, can't carry on the effective fusion.

Dempster's rule of combination has some defects and limitations in its application. When there is no conflict or low conflict between evidences, the reasoning of evidences is basically normal, but when there is serious conflict between evidences, the combination result is often inconsistent with the actual situation, that is, the combination rule cannot handle the conflict [19]. Entropy theory holds that the smaller the entropy of information is, the greater the utility value of information will be, and the greater the weight of indicators will be. In this paper, according to the entropy theory, the information entropy of evidence is dynamically calculated to form a correction coefficient, which is used to modify the basic probability distribution (BPA), improve Dempster's rule of combination, and reduce the degree of evidence conflict. Specific improvements are as follows:

If m evaluation objects and n attributes constitute a cube $\mathrm{D} = \left(a_{ij}\right)_{m \times n}$, then the information entropy of the jth attribute is [20]

$$H_j = -\mathrm{k} \sum_{i=1}^{m} f_{ij} \ln f_{ij}$$

where $f_{ij} = {}^{a_{ij}} / \sum_{i=1}^{m} a_{ij}^{\mathrm{k}} = 1/\ln \mathrm{m}^{\text{if}}\ f_{ij} = 0, f_{ij} \ln f_{ij} = 0$.

Data set $\mathrm{D} = \left(a_{ij}\right)_{m \times n}$, then at time t, the index weight of the jth attribute is

$$\omega_t(j) = \frac{1 - H_j}{n - \sum_{j=1}^{n} H_j} \tag{8}$$

where $0 \le \omega_t(j) \le 1$, $\sum_{j=1}^{m} \omega_t(j) = 1$, When a new evaluation object is added or exited, the index weight of the attribute is recalculated. According to the importance weight of indicators, the weight of evidence under the recognition framework is adjusted to obtain

$$U\left(m_j\right) = m_j \omega_t(j) \tag{9}$$

According to the information entropy, the weight of evidence is adjusted, and the original basic probability distribution is modified, that is, the new probability distribution function (BPA) can be obtained

$$m_i'(A) = \begin{cases} U\left(m_i\right) m_i(A), & A \neq \Theta \\ 1 - \sum_{B \subseteq \Theta} U\left(m_i\right) m_i(B) & , A = \Theta \end{cases} \tag{10}$$

After revising the basic credible number of evidence according to the weight, Dempster's rule of combination can distinguish the importance of evidence, so as to effectively alleviate the conflict caused by different importance of evidence in the process of evidence combination.

Definition 6: class of probability functions.

$$f(A) = \mathrm{Bel}(A) + \frac{|A|}{|\Theta|} \times (Pl(A) - Bel(A)) \tag{11}$$

Where $|A|$ and $|\Theta|$ are the number of elements in A and Θ respectively. The certainty of A proposition can be expressed by the quasi-probability function.

3 Trust Evaluation Model for Big Data

3.1 Distributed Collaborative Trust Management Framework

The credibility of big data largely depends on the credibility of the data source that publishes the data, which is determined by the direct and indirect trust of the data source. Based on the existing big data production system, this paper proposes a distributed collaborative trust management framework, which collects, stores and computes the direct trust, indirect trust and comprehensive trust of big data sources through distributed trust agents. Direct trust can be obtained by improving D-S evidence theory, by calculating the objective data of the big data system and the subjective data of user evaluation, and indirect trust can be obtained by calculating the reputation value of the data source through the big data network. Distributed trust agents not only store and manage the trust values of all data sources of the big data provider where they are located, but also store the trust values of adjacent nodes for other distributed agents to query. The distributed collaborative trust management framework is shown in the Figure 2.

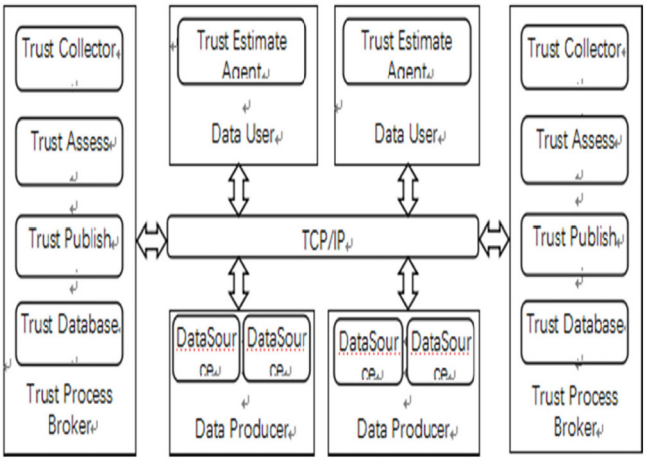


Figure 2. Distributed collaborative trust management framework

The main actors of distributed collaborative trust management framework are :(1) big Data Producter (DP), which provides source data services to customers for profit. In the specific environment of big data, DP mainly provides source materials to data manufacturers in the form of data files and API. (2) Big Data Source (DS), providing text, video, pictures and other types of Data. (3) Big Data User (DU), using the Data provided by DP. In addition, it can request the trust value of DP from TPB, so as to select the most trusted DS. (4) Trust Process Broker (TPB), which provides the credibility information of DS and is responsible for publishing, updating and sharing the global and local trust values of DS. (4) Trust Estimate Agent (TEA), who is responsible for evaluating the subjective trust value of DS and querying the trust value of DS. (5) Trust DataBase, TDB, which stores trust values of DS nodes directly managed by TPB and link information of neighboring TPB nodes.

TPB is distributed in the Internet and entrusted by different big data provider DS to provide information services related to trust computing. When DU requests a big data source with trust requirements, TPB is called. A TPB can represent multiple DS under a big data application, and a DS can also be represented by multiple TPB. TPB obtained DU's subjective evaluation information of DS from TEA, and further calculated the objective trust value of DS. TPB can interact with each other, exchange and spread trust information. DU can obtain the letter evaluation of DS from multiple TPB.

TPB mainly records and maintains the following information:

(1) DS set represented by the TPB;

(2) the neighbor TPB set trusted by the TPB.

(3) The trust calculation algorithm used by the TPB to evaluate the trust value of DS and other TPB;

(4) the trust strategy used by the TPB in different big data environments.

(5) Data processing and trust computing module.

3.2 Trust Measurement

3.2.1 Direct Credibility Calculation

Direct credibility is expressed by T_d, which reflects TPB's direct trust to data sources. The direct trust calculation in this paper considers the multi-dimensional attribute evidence that affects the trust of big data online data sources. In the trust model of this paper, Θ is defined as $(T,-T)$ and the relationship between TPB and DS is divided into trust (T), distrust $(-T)$, uncertainty $(T,-T)$. According to the data characteristics of the large data, this article selects the data source performance E1, the data source properties normative E2, data item null frequency E3 and E4 data arrival rate **[11]**, the data source properties by the TPB through big data platform of CPU, memory, disk I/O, load, objective data technologies such as data source, data standardization, data items, frequency of null values, data arrival rate by DU subjective evaluation.

(1) DS basic probability distribution. The basic probability distribution is shown in the Table 1.

Table 1. Basic probability distribution table

	E1	E2	E3	E4
{T}	m_{11}	m_{21}	m_{31}	m_{41}
{-T}	m_{12}	m_{22}	m_{32}	m_{42}
{T,-T}	m_{13}	m_{23}	m_{33}	m_{43}

The basic probability distribution functions of E1,E2,E3 and E4 in the identification framework Θ are m_1, m_2, m_3, m_4 respectively.

(2) Based on the basic probability distribution corresponding to each evidence, the importance weight of evidence is calculated by using equations (1)~(7)

$$U\left(m_j\right), \quad j=1,2,3,4 \tag{12}$$

(3) The modified BPA function can be obtained by substituting the importance weight into equation (10),

$$\left\{m_i'(T), m_i'(-T), m_i'(T,-T)\right\}(i=1,2,3,4) \tag{13}$$

(4) Finally, according to formula (4), attribute evidence is synthesized to obtain trust triples $\left\{m_i'(T), m_i'(-T), m_i'(T,-T)\right\}$ representing direct trust of big data sources.

(5) Formula (11) class probability function is used to calculate the direct trust value. As a measure of the imprecision of trust, the reliability space of uncertain events is divided according to the probability of occurrence of trusted and untrusted events in the data source, and the two events are assigned to be trusted and untrusted.

Define 9 TPB's direct trust in DS

$$T_d = f(T) = Bel(T) + \frac{|\{T\}|}{|\{T,-T\}|} \times (Pl(T) - Bel(T)) \tag{14}$$

$|\{T\}|=1, |\{T,-T\}|=2$. Formula (14) comprehensively considers the trusted part, the untrusted part and the uncertain part of the data source trust relationship. This method can accurately evaluate the direct trust relationship between big data sources and users, is more intuitive and closer to the authenticity of trust. TPB's direct trust in DS is stored in TPB's trust value database.

3.2.2 Indirect Reliability Calculation

In the big data production environment, some big data users cannot obtain the trusted value of DS directly through the TPB they are connected to, and they need to obtain the trusted value of the target data source indirectly through other TPB. Therefore, it is necessary to combine the recommendation information of TPB to calculate the trusted value of DS.

Define 10 recommendation trust. The trust judgment made by TPB based on the evaluation of DS provided by the third-party user TPB is called recommendation trust and also called indirect trust.

In this paper, PageRank algorithm is used to calculate the recommended trust value. According to the principle of sociology, high-quality data sources provided by well-known institutions will be more accepted and recognized by users, and will be used more often. Therefore, we introduce the PR value of data source node as the recommendation trust of data source. At time t, the recommendation trust formula of data source A is as follows:

$$T_r(t) = \frac{1-a}{N} + a \sum_{T_i \in S(A)} \frac{PR\left(T_i\right)}{C\left(T_i\right)} \tag{15}$$

Where a is the empirical value that the data source jumps to other data sources with a certain probability, 0.85 is adopted in this paper, and N is the total number of all data sources in the big data production network at time t. The recommendation trust degree of data source is composed of two parts: the data source PR value corresponding to the data source served by the source data source, namely a sigma $a \sum_{T_i \in S(A)} \frac{PR\left(T_i\right)}{C\left(T_i\right)}$; The PR value contributed by one data source randomly selected by the user from many data sources, $\frac{1-a}{N}$.

3.2.3 Trust Propagation and Integrated Credibility Calculation

The trust propagation network of TPB is formed by the trust relationship between each node and its neighbors. With this network, you can get a trust value for one from somewhere else. A TPB trust propagation network is shown below. The solid line between two nodes indicates mutual trust between nodes.

In the Figure 3,we suppose that DU_B requests the trust value of DS_E from TPB_B, and there is no value about DS_E on TPB_B, so TPB_B sends the request to its neighbor TPB_C, then propagates the request to TPB_D, and finally finds the trust record of DS_E on TPB_E. The propagation of the trust message continues until all connected nodes in the network are covered. In the example, we assume that the requested message will be received from all of its neighbors through the following paths,TPB_B->TPB_C->TPB_D->TPB_E and TPB_B->TPB_C ->TPB_E, as long as the trust information for the data source DS_E being asked is relevant. For each path, TPB_E will reply the requested DS_E trust value along the opposite direction of the path. Since the attenuation of trust is universal in trust propagation, the longer the path, the more severe the attenuation of trust, and the higher the transmission cost and risk of data. Therefore, when there are multiple paths in source TPB and target TPB, we choose the shorter path as the trust propagation path. Compared to indirectly recommended entities, entities tend to trust the entities they believe in and directly recommend the entities they believe in, so shorter paths show higher trust than longer paths. Comprehensive trust is defined as

$$T(t) = T_d(t) \otimes T_{r1}(t) \otimes T_{r1}(t) \otimes \ldots \otimes T_{rn}(t) \tag{16}$$

where n is the number of TPB between DS and DU.

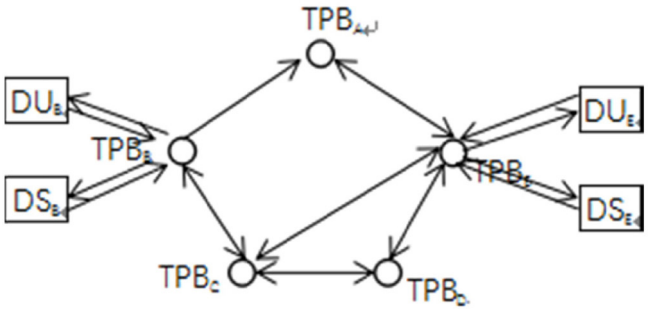


Figure 3. Trust propagation network

4 Experiments

In order to evaluate the trust model proposed in this paper, we use Netlogo to simulate a big data network composed of 100 DS, including multiple DS and a large number of DU. The reliability of DS is measured by trusted value, untrusted value and uncertain value. A total of 100 DU were simulated in the experiment, and attribute data were generated randomly. The experimental steps are as follows:

(1) Simulate the evaluation data of 100 DU for DS;

(2) Use information entropy to calculate the variable weight of the four indicators, and generate the basic probability assignment function by weighting;

(3) Use the improved Dempster combination rule to calculate DS direct trust value;

(4) Use PageRank to calculate DS indirect trust value;

(5) Calculate the comprehensive trust value.

4.1 Direct Trust Calculation

In this paper, TPB uses four indicators to evaluate the reliability of big data DS, where E1 represents the performance of data sources and is an objective indicator, which is calculated according to the CPU, memory and load of DS; E2 refers to data normalization of data sources; E3 to null frequency of data items; E4 to data arrival rate, all of which are subjective indicators generated by DU evaluation. 10 nodes are selected and their scores are shown in the Table 2.

Table 2. Big data source attribute value

NO	E1	E2	E3	E4
1	0.85	0.82	0.14	0.85
11	0.81	0.84	0.03	0.67
21	0.62	0.21	0.53	0.77
31	0.79	0.19	0.24	0.39
41	0.85	0.41	0.06	0.81
51	0.23	0.78	0.77	0.64
61	0.41	0.24	0.11	0.12
71	0.98	0.89	0.24	0.43
81	0.25	0.53	0.05	0.17
91	0.39	0.37	0.51	0.71

According to the weight calculation in formula (8), the variable weight of each evaluation index can be obtained. The results can be seen in the Table 3.

Table 3. The variable weight of each evaluation index

E1	E2	E3	E4
0.135728	0.197911	0.368532	0.297829

According to formula (1)~(11), the above 10 data source nodes have direct trusted values, as shown in the Table 4.

Table 4. Direct credibility calculation

NO	$m_i(T)$	$m_i(-T)$	$m_i(T,-T)$	$Bel(T)$	$Pl(T)$	$f(T)$	T_d
1	0.253398	0.048201	0.698401	0.253398	0.746602	0.626699	0.626699
11	0.252099	0.049252	0.698651	0.252099	0.747901	0.62605	0.62605
21	0.047832	0.16187	0.790299	0.047832	0.952168	0.523916	0.523916
31	0.124814	0.169068	0.706117	0.124814	0.875186	0.562407	0.562407
41	0.122186	0.027909	0.849905	0.122186	0.877814	0.561093	0.561093
51	0.166183	0.128251	0.705565	0.166183	0.833817	0.583092	0.583092
61	0.052171	0.15458	0.79325	0.052171	0.947829	0.526086	0.526086
71	0.283307	0.021123	0.69557	0.283307	0.716693	0.641654	0.641654
81	0.036412	0.104102	0.859487	0.036412	0.963588	0.518206	0.518206
91	0.020045	0.020045	0.95991	0.020045	0.979955	0.510022	0.510022

4.2 Indirect Trust Calculation

The big data network of 100 nodes was generated through Netlogo simulation, as shown in the Figure 4. 10 nodes were selected to calculate PageRank, as shown in the Table 5.

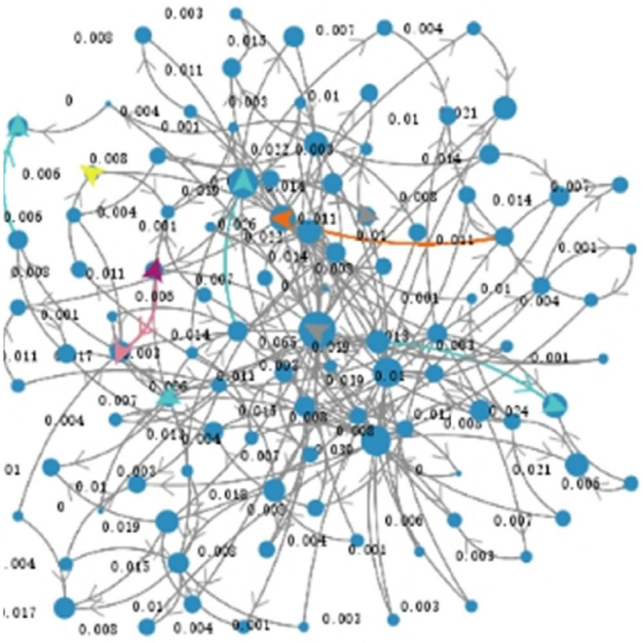


Figure 4. TPB trust propagation network

Table 5. PageRank calculation

who	rank	is	who	rank	is	who	rank	is
0	0.026389	19	34	0.001389	1	67	0.009722	7
1	0.018056	13	35	0.0125	9	68	0.002778	2
2	0.065278	47	36	0.005556	4	69	0.016667	12
3	0.038889	28	37	0.016667	12	70	0.004167	3
4	0.004167	3	38	0.009722	7	71	0.006944	5
5	0.038889	28	39	0.011111	8	72	0.001389	1
6	0.022222	16	40	0.001389	1	73	0.008333	6
7	0.0125	9	41	0.038889	28	74	0.006944	5
8	0.013889	10	42	0.004167	3	75	0.001389	1
9	0.004167	3	43	0.008333	6	76	0.002778	2
10	0.005556	4	44	0.011111	8	77	0	0
11	0.008333	6	45	0.0125	9	78	0.004167	3
12	0.006944	5	46	0.005556	4	79	0.001389	1
13	0.020833	15	47	0.006944	5	80	0.005556	4
14	0.015278	11	48	0.004167	3	81	0.002778	2
15	0.015278	11	49	0.013889	10	82	0.004167	3
16	0.013889	10	50	0.009722	7	83	0.006944	5
17	0.009722	7	51	0.002778	2	84	0.001389	1
18	0.013889	10	52	0.006944	5	85	0	0
19	0.020833	15	53	0.005556	4	86	0.001389	1
20	0.013889	10	54	0.008333	6	87	0	0
21	0.009722	7	55	0	0	88	0.002778	2
22	0.011111	8	56	0.005556	4	89	0.002778	2
23	0.022222	16	57	0.023611	17	90	0.008333	6
24	0.008333	6	58	0.002778	2	91	0.008333	6
25	0.009722	7	59	0.009722	7	92	0.011111	8
26	0.009722	7	60	0.001389	60	93	0.006944	5
27	0.008333	6	61	0.002778	61	94	0.004167	3
28	0.019444	14	57	0.023611	17	95	0.002778	2
29	0.008333	6	62	0.011111	8	96	0.008333	6
30	0.011111	8	63	0.004167	3	97	0.0125	9
31	0.019444	14	64	0.001389	1	98	0.004167	3
32	0.008333	6	65	0.002778	2	99	0.015278	11
33	0.016667	12	66	0.009722	7			

According to formula (15), the Table 5 can be obtained.

Construct DU1 and query the trust degree of DU91 through TBP, then

$$T(t) = T(t) = T_d(t) \otimes T_r(t)$$

$$= 0.319459 * 0.00833 = 0.00266$$

It can be seen from the experimental results that the reliability has relatively higher trusted value and relatively lower untrusted value and uncertain value. Untrustworthiness has relatively higher untrustworthiness value and relatively lower trustworthiness value and uncertainty value; And random has relatively higher uncertain value and relatively lower trusted value and untrusted value.

5 Conclusion

This paper proposes a multi-attribute decision trust assessment model and distributed collaborative management framework based on improved D-S evidence theory in the context of big data and evaluates the credibility of big data sources through subjective and objective evidence. In this framework, TPB uses D-S theory to collect, process and evaluate the trust information in the big data environment, and can obtain the direct trust value of DS from a single DU perspective. In order to share in the big data environment of multiple data source provider of trust information, this model is to establish trust between the TPB and TPB transmission network, when a TPB received about a trust information request, will spread through trust network to forward the request to neighbors, trust network spread through every trust relation with its neighbors and form, using PageRank from the angle of all the assembled by indirect trust value. Using direct and indirect trust values, we can calculate the comprehensive trust value of the data source. Experiments show that our proposed trust management framework is effective and robust in identifying the trusted and the untrusted.

Acknowledgements

This work is supported by the National Quality Infrastructure project that is Key R&D Program of China under Grant 2018YFF0212106.

This work was supported in part by National Natural Science Foundation of China “Theory and Key Technologies of Distributed Autonomous Security Supervision for Data Opening and Sharing” NO. 61962009, 2020/01-2023/12.

This work is supported by Major Scientific and Technological Special Project of Guizhou Province 20183001.

This work is supported by Foundation of Guizhou Provincial Key Laboratory of Public Big Data No.2018BDKFJJ005&No.2018BDKFJJ013

This work is supported by Talent project of Guizhou Big Data Academy. Guizhou Provincial Key Laboratory of Public Big Data [2018] 01.

References

[1] Y. Gao, *The Research and Implementation of Credibility Assessment of E-Commerce Service Based on Third-Party*, M. S. Thesis, Beijing University of Post and Telecommunications, Beijing, China, 2015.

[2] M. Kosfeld, Trust in the Brain. Neurobiological Determinants of Human Social Behavior, *EMBO Reports*, Vol. 8, No. Suppl 1, pp. S44-S47, July, 2007.

[3] A. Abdul-Rahman, S. Hailes, Supporting Trust in Virtual

Communities, *33rd Annual Hawaii International Conference on System Sciences*, Maui, Hawaii, 2000, pp. 1-9.

[4] V. Buskens, The Social Structure of Trust, *Social Networks*, Vol. 20, No. 3, pp. 265-289, July, 1998.

[5] Y. Wang, J. Vassileva, Trust and Reputation Model in Peer-to-Peer Networks, *Third International Conference on Peer-to-Peer Computing (P2P2003)*, Linkoping, Sweden, 2003, pp. 150-157.

[6] B. Gupta, H. Kaur, Namita, P. Bedi, Trust Based Access Control for Grid Resources, *International Conference on Communication Systems and Network Technologies*, Jammu, India, 2011, pp. 678-682.

[7] Y.-L. Sun, W. Yu, Z. Han, K. J. R. Liu, Information Theoretic Framework of Trust Modeling and Evaluation for Ad hoc Networks, *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 2, pp. 305-317, February, 2006.

[8] R.-J. Feng, X.-F. Xu, X. Zhou, J.-W. Wan, A Trust Evaluation Algorithm for Wireless Sensor Networks Based on Node Behaviors and D-S Evidence Theory, *Sensors*, Vol. 11, No. 2, pp. 1345-1360, January, 2011.

[9] A. Josang, A Logic for Uncertain Probabilities, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 9, No. 3, pp. 279-311, June, 2001.

[10] L. Mui, M. Mohtashemi, A. Halberstadt, A Computational Model of Trust and Reputation, *35th Hawaii International Conference on System Sciences*, Big Island, HI, 2002, pp. 2431- 2439.

[11] X. Wu, Y. Wang, Y. Yuan, Multi-dimensional Trust Evaluation Method Based on D-S Evidence Theory, *Computer and Digital Engineering*, Vol. 47, No. 2, pp. 367-372, February, 2019.

[12] J.-H. Li, H. Wang, J.-J. Lei, Trust Evaluation Model Based on Entropy Weight Method in Trusted Network, *Journal of Central China Normal University (Natural Sciences)*, Vol. 53, No. 1, pp. 26-29, February, 2019.

[13] J. Sänger, C. Richthammer, S. Hassan, G. Pernul, Trust and Big Data: A Roadmap for Research, *2014 25th International Workshop on Database and Expert Systems Applications*, Munich, Germany, 2014, pp. 278-282.

[14] M. Albanese, Measuring Trust in Big Data, *International Conference on Algorithms and Architectures for Parallel Processing*, Vietri sul Mare, Italy, 2013, pp. 241-248.

[15] G. Li, T.-Q. Li, X.-R. Cheng, H.-Y. Wang, Credibility Measurement Method of Big Data, *Computer Engineering and Design*, Vol. 38, No. 3, pp. 652-658, March, 2017.

[16] A. P. Dempster, Upper and Lower Probabilities Induced by a Multivalued Mapping, *The Annals of Mathematical Statistics*, Vol. 38, No. 2, pp. 325-339, April, 1967.

[17] A. P. Dempster, A Generalization of Bayesian Inference, *Journal of the Royal Statistical Society*: *Series B*, Vol. 30, No. 2, pp. 205-247, July, 1968.

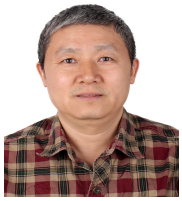
[18] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[19] X. Guan, X. Yi, X.-M. Sun, Y. He, Efficient Fusion Approach for Conflicting Evidence, *Journal of Tsinghua University*, Vol. 49, No. 1, pp. 138-141, January, 2009.

[20] Y. Mei, J.-W. Ye, Z.-G. Zeng, Entropy-weighted ANP Fuzzy Comprehensive Evaluation of Interim Product Production Schemes in One-of-a-kind Production, *Computers and Industrial Engineering*, Vol. 100, pp. 144-152, October, 2016.

Biographies

Yu-Ling Chen, female, born in 1983, associate professor, works in State Key Laboratory of Public Big Data of Guizhou University, mainly studies information security, block chain, privacy protection of big data.



Yu-Jun Liu, male, born in 1963, senior engineer, works in Technical Center of Beijing Customs, mainly studies EMC of Electronic products and information security.

Wei-Fa Zheng, male, born in 1980, senior engineer, works in the Network Information center of Guangdong University of Finance and Economics, mainly studies signal processing and information security.

Jing-Yao Chen, female, born in 1996, postgraduate student, studies in the electronics department of shantou university.