Advances in Social Science, Education and Humanities Research, volume 505
6th International Conference on Social Science and Higher Education (ICSSHE 2020)

Trust Evaluation of Users in Social Community Based on Virtual Identity Association Analysis

Jianying Xiong[1,*] and Juan Luo[1]

[1] *Security Management Department of JiangXi Police College, Nanchang, Jiangxi 330013, China*
**Corresponding author. Email: special8212@sohu.com*

ABSTRACT
Effective network trust evaluation can promote the self-discipline of users' information behavior. Due to the data source of traditional trust assessment is relatively single, the purpose of this study is to obtain the comprehensive performance of the same entity in different network communities to make a comprehensive evaluation of users' trust. The work designed an association verification by integrating heterogeneous data and crawling Internet public information, and proposed a user trust evaluation model for social network users based on multi-community virtual identity trust fusion. The case shows the feasibility and effectiveness of the method, which is conducive to the online virtual identity verification and tracking, and promotes social behavior self-discipline.
Keywords*: Social network governance, trust evaluation, trust fusion, multi-platform virtual identity*

1. INTRODUCTION

Social network has become an important channel for people to obtain or share information. It allows everyone to participate in information release and dissemination. This barrier-free interactive mode not only provides great convenience for people's information sharing, but also brings some prominent problems. Especially the rapid spread of risk information such as Internet rumors, politically sensitive information, inciting online public opinion, false advertising information and so on. In social network information dissemination, users become gatekeepers. Due to the restriction of culture, education and other factors, the huge group of social network users show uneven state, resulting in the serious weakening of the gatekeeper of information sources, thus weakening the role of gatekeeper. In social networks, many problems of information dissemination are often caused by user nodes, so it is of great practical significance to study the credibility of users.

Social activities on the Internet are frequent. People register different IDs on various platforms to carry out network social activities, which form a new identity - internet virtual identity. A large number of virtual identity communication forms the social network diagram. Because the network virtual identity does not need real-name authentication, this increases network security maintenance and supervision. Traditional trust evaluation mainly relies on users' disclosure of information and behavior information in the social network. In the context of big data, it is not enough to use the behavior of users on a certain platform to evaluate their trust. We need to integrate multiple virtual communities, share the data behavior of virtual users, and conduct multi-dimensional cross validation and evaluation of user trust. This multi-platform Association fusion method is not only conducive to the network virtual identity verification and tracking, but also can promote the self-discipline of users' social behavior and purify the network at the source.

2. RELATED WORK

Dishonesty and malicious behavior of users are inevitable in a complex social network environment. Designing network application algorithm is also an important aspect of social network governance. Trust mechanism is an essential solution to ensure the security of social networks. It also can be understood as one party's recognition of the other party's reliability, which reflects a mutual trust relationship [2]. Roy constructs a social media trust scoring algorithm for the problem of social network advertising communication, which is used to strategy the trust level of individual users in social network, and adjusts the trust score according to the network decision-making factors in different situations; Cheng studies the influencing factors of trust in interpersonal communication, group communication and mass communication in social networks, and verifies trust factors and their relationship through interviews with 115 participants in WeChat [4]. In the study of user behavior, relationship, and other attribute characteristics, information behavior is used as the information source to evaluate trust. The network information behavior is also a kind of social behavior. The users play different roles in the network society, and the information interaction is still the interaction between people. Publishing and spreading false information, bad information, uncivilized comments, malicious slander and vicious language attacks in social networks are easily spread by network groups. The convenience of content generation in social network requires users to self-judge and correct information behavior [5]. Nie uses sociology,

psychology and behaviourism to study user information behavior, reveals users' internal personality traits, motivation, cognition, etc., and proposes that trust, culture and other external social attributes have a significant impact on information behavior [6].

Virtual identity association is a new direction in the field of social network analysis in recent years. In particular, virtual identity association has great commercial potential and practical challenges, which has attracted more and more attention from domestic and foreign professional scholars. Zafarani R and Liu proved the method of cross social network platform virtual identity Association [7]. They used behavior modeling method to construct and evaluate user characteristics, used the minimum information contained in the user's registered user name in the social network platform to mine a large amount of information, and finally carried out effective virtual identity among multiple social network platforms through supervised learning algorithm relation. Siyuan Liu classifies user registration information in social network platform into text information and visual information [8]. Text information includes user name, gender, label, etc., and visual information refers to the user's head image in user information. Andrenunes and Pavel calado [9] construct the user's feature vector through the following three types of features: similarity of user's personal information (user name, user's gender, user's email), similarity of user's behavior information (tag attribute of user's published article) and similarity of user's Internet information (user's concerned friends). By considering the similarity of features of users in different platforms, whether different virtual identities belong to the same real person is identified, Scholars have studied the collection, merging and storage of virtual identity data, and then verified it by building a prototype system of virtual identity knowledge map [10]. How to extract the relationship between two nodes in social network, some scholars also proposed a method. This method mainly divides the virtual identity graph into separate communities by using the community discovery algorithm copra and GraphX computing framework, and realizes the common friend calculation algorithm of two nodes to mine the relationship between the two nodes [11]. Based on the existing research foundation, through the integration and processing of heterogeneous data, this paper realizes the data resource integration scheme in the case of decentralized platform and multi-database data sources. Then proposes a technology of virtual and real identity chaining, association and analysis based on big data to comprehensively evaluate and verify the trust of social network users.

3. CONSTRUCTION OF TRUST MODEL

3.1. Model Framework

Virtual identity information is the identity information of someone or an organization in cyberspace, such as user A uses mailbox in a forum 123@aa.com Registered account number 123456, nicknamed ABC. All the above email 123@aa.com, account 123456, nickname ABC information is considered as virtual identity information. At present, due to the various network virtual accounts established by individuals or organizations, each person or organization may have multiple virtual accounts on one or more platforms. The framework of the multi-platform Association fusion trust evaluation method proposed in this paper is shown in Figure 1. The main contents include: obtaining the virtual identity information of multiple platforms; analyzing the virtual identity information and extracting metadata; associating the metadata; identifying the virtual account according to the same identity; information fusion and comprehensive trust evaluation for all platform virtual accounts. At the same time, the method can determine the virtual identity to a certain extent, which provides a reliable scheme for tracing the virtual identity.

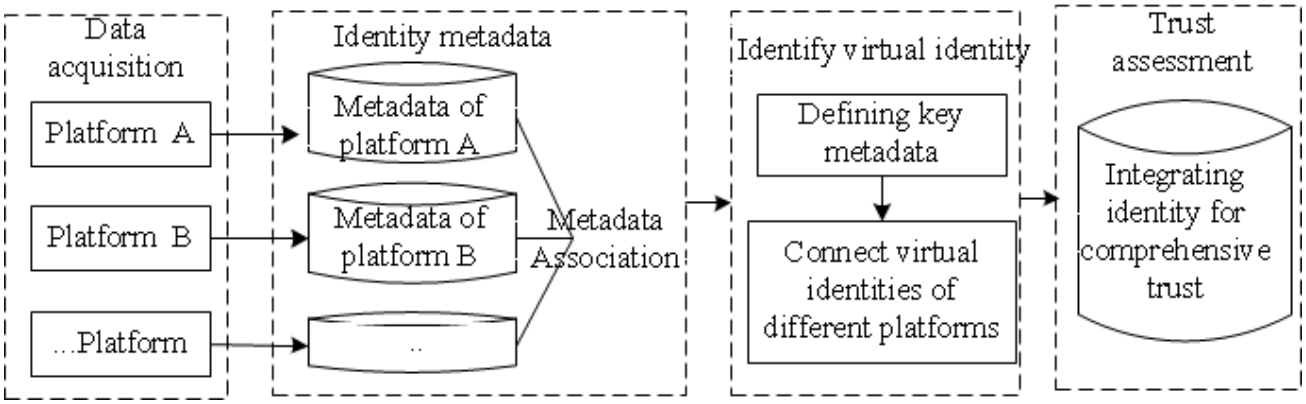

Figure 1. The overall framework of the model

3.2. Trust Fusion Process

3.2.1. Acquisition of virtual identity information

The relevant data of virtual account can be obtained by means of crawler and interface transmission. Crawler technology, namely web crawler, is a program or script that automatically grabs the information of the world wide web according to certain rules. Figure 2 provides an example of obtaining virtual identity information from the relevant web pages of X website. The user meta information includes nickname, gender, date of birth, email address, occupation, grade, etc.; however, some sensitive information is transmitted through the interface, such as bound mobile phone number, real name, ID card, micro signal, etc. The sharing of sensitive data is based on the formation of a certain alliance among various platforms. In order to govern data sharing in cyberspace, this part of information is not transmitted in the form of plaintext, but encrypted with a unified public key. This can not only protect the privacy of users, but also effectively connect the identity of virtual identity.

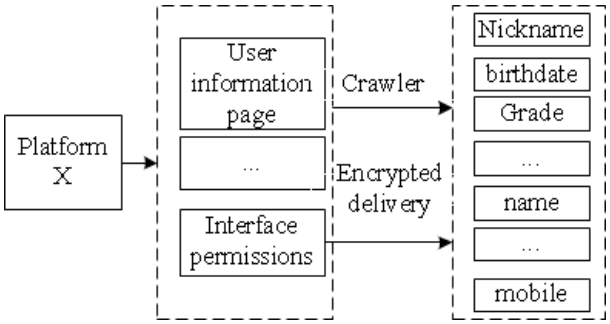

Figure 2. Data acquisition

3.2.2. Metadata extraction and Association

After obtaining the virtual identity information, it is necessary to analyze the virtual identity information. The information obtained by the crawler needs to be pre-processed to obtain the corresponding virtual identity data. If the transmission is based on the authorization of the interface, the traffic needs to be restored to obtain the corresponding data. The above extracted data contains the attribute and attribute values. We define "property value" as metadata. For example, "name Zhang San" is an attribute, and "Zhang San" is the value of the corresponding "name" attribute. We define "name Zhang San" as metadata. Since each virtual identity has a fixed account, we choose the account ID as the unique ID of the virtual identity information of the account in a certain platform x, which is recorded as (account ID: information source x). Therefore, the obtained virtual identity information at least includes account ID. In addition, it also includes some other virtual identity information, such as email, telephone number, QQ number, password, nickname, etc. The more information extracted, the more perfect the virtual identity information association network is constructed. If the website information is incomplete, the information that cannot be extracted will be set to null value. The authenticity of the name, ID card number, mobile phone number, etc. obtained by the interface is high. It does not rule out the errors caused by borrowing other people's information, using expired information, and lax website verification. We take the above information as the main series clue. If the information is inconsistent, the similar probability will be returned according to other information. Considering the convenience of subsequent calls, we save them to the corresponding database in the form of triples (account ID: information source) - attribute value.

3.2.3. Virtual identity

Based on the extracted metadata, the association network is formed. As shown in Figure 3, the tree association network can more intuitively show the association between account ID, attribute and attribute value. The attribute with high value for identity is set as the key metadata, which is generally the encrypted privacy data obtained by the interface, such as mobile phone, name, WeChat and other data in the figure. The tag format of metadata is ((account ID, dependency coefficient), source), and the initial dependency coefficient of the tag is 1. For example, the initial tag of the metadata "name Zhang San" from platform x with account ID of zhangsan0123 is ((zhangsan0123, 1), x). The metadata from different sources and the initial tags established in the previous step are stored in a unified way. In this process, since the final formation is the fusion of virtual identity information, we do not need to consider the source of information. So we can remove the information source in the label. That is, for metadata ((zhangsan0123: x) - mailbox in source X - zhangsan1990@163.com) After removing the source, the label becomes (zhangsan0123,1).

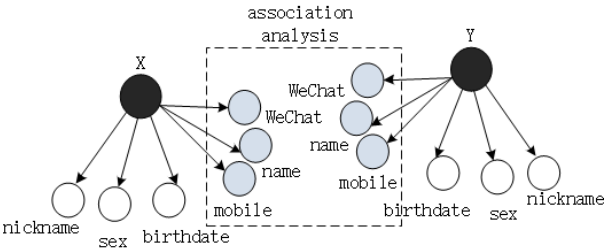

Figure 3. User Metadata

3.2.4. Trust fusion

For different platform dimensions, comprehensive trust can be calculated from the quality of platform information sources and the selection of trust characteristics.

3.2.4.1. Direct platform trust fusion.

According to the quality of platform information source, the weights of different platforms can be set based on fuzzy theory, and the trust synthesis can be carried out after calculating the weights of each attribute. Trust fusion formula is shown in (1), T is comprehensive trust, (w_1, $w_{2...}w_n$) is the weight of different platforms, (t_1, t_2… t_n) is the trust measure of different platforms. The user trust of each platform can be obtained by the evaluation system of different platforms, such as normalization of user level index.

$$T = \sum_{i=1}^{n}(w_i * t_i) \quad (1)$$

For the calculation of the weight of each platform, if the platform has strong generality, it means that the network users have better recognition. The social users are more likely to form network connection with other people, and the platform has more authority on the user's trusted state. Therefore, according to the universality of the platform, the platform is set to different levels as Table 1, and finally the standardized value of each grade is taken as the weight.

Table 1. Platform authority level

	High	Medium	low
Authority value (d)	3	2	1

$$w_i = d_i / \sum_{i=1}^{n} d_i \quad (2)$$

3.2.4.2. Trust feature fusion method.

According to the trust elements needed in trust evaluation theory, this method constructs the required complete trust elements (Mt_1, Mt_2,…Mt_n) from two aspects of direct trust and indirect trust. Extract and merge relevant metadata from all platforms as data sources. Direct trust takes information disclosure of information source as node identity trust; indirect trust takes information interaction behavior between people as node behavior trust; feedback behavior generated by information interaction of various platforms as trust reward and punishment adjustment mechanism to generate node reputation comprehensively.

In the metadata fusion of the trust elements, the features provided by the authoritative platform can be preferentially selected, but the behavior punishment of each platform will be included in the trust evaluation with the same weight. The fused trust feature metadata will not indicate the information source and form a triplet: (attribute, value, weight). The fusion calculation method is shown in (3), where T is the comprehensive trust, w_d is the direct trust DT weight, w_r is the indirect trust RT weight value, and r is the penalty coefficient. For the calculation of direct trust and indirect trust, there are many calculation methods in the existing research, this paper does not focus on the model, here we can directly use the weighted method.

$$T = (1 - r) * (w_d * DT + w_r * RT) \quad (3)$$

4. MODEL EXAMPLE

4.1. Scene Description

The platform can directly provide an interface for users to search. Users only need to select one integrated platform and input the virtual account directly, as shown in figure 4. The search process mainly includes the following steps: receiving the query statements input by users; forming a query tree based on the syntax rules of keywords; after being read into memory, the query tree is used to search the index to get the results; the search results are sorted according to the relevance of the query, and the query results are returned to the user.

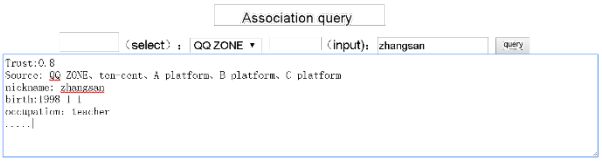

Figure 4. Virtual identity Association query

4.2. Trust Computing Example

For example, enter "Zhangsan" and select the original source platform "QQ zone". The system will search from the metadata stored in all integrated platforms. For example, a metadata ((zhangsan1234: x) - mailbox in source a -zhangsan1910@163.com) The initial label is established ((zhangsan1234, 1), a);. For another metadata in source B ((zhangsi0123: b) - mailbox-zhangsan1910@163.com) The initial label established is ((zhangsi0123, 1), b); for another metadata in source c ((zhangsi0234: C) - mailbox -zhangsan1990@163.com) The initial label established is ((zhangsi0234,1), c). In this way, the account "Zhangsan" can be associated with three other platforms (a, B, c), and their behavior can be further tracked.

For the above three metadata, their "attribute value" is "mailbox" -zhangsan1990@163.com ”However, due to different sources or different account IDS, three initial tags have been established for their three metadata. However, since the "attribute value" is exactly the same, the sum of their dependency coefficients must be 1. Therefore, the dependency coefficients of the above three metadata are renewed When the metadata related to trust characteristics is merged, it needs to be weighted according to the authority of the platform. Suppose the authority level of three platforms A, B and C is 3,2,1; the user level of the three platforms is evaluated as (A: 0.9, B: 0.8, C: 1), then the comprehensive trust is: T = 0.9 * 3 / (3 + 2 + 1) + 0.8 * 2 / (3 + 2 + 1) + 1 * 1 / (3 + 2 + 1) = 0.89.

Suppose that the direct trust, indirect trust and penalty coefficient of the three platforms are (A: (0.9,0.7,0.1), B:

(0.8,0.6,0.2), C: (0.7,0.9,0.1)). The weights of direct trust and indirect trust are {0.6,0.4}. The integrated trust calculation process is as follows:
1) direct trust DT = 0.9 * 3 / (3 + 2 + 1) + 0.8 * 2 / (3 + 2 + 1) + 0.7 * 1 / (3 + 2 + 1) = 0.84;
2) calculate indirect trust RT = 0.7 * 3 / (3 + 2 + 1) + 0.6 * 2 / (3 + 2 + 1) + 0.9 * 1 / (3 + 2 + 1) = 0.7;
3) Comprehensive trust t calculation: T = (1-0.1) * (0.84 * 0.6 + 0.7 * 0.4) = 0.7.
The second method will consume more resources than the first method, but it can unify the evaluation criteria of trust in different platforms and strengthen the role of the penalty coefficient.

5. CONCLUSION

When facing a large number of virtual network platforms, it is not enough to rely on a single platform data in the process of virtual community governance and virtual identity tracking analysis. We also need to integrate multiple platforms for association analysis, so we need to integrate multi-platform and multi-source data, and reasonably use the data for retrieval and analysis, and show the results. In this paper, through the analysis of the characteristics of network virtual identity, the identification and association analysis model of virtual identity is established. The model extracts the virtual identity metadata of each platform, uses the key metadata for multi-platform series analysis, and weights the user trust according to the universality characteristics of each platform. Due to the diversity of data sources, in addition to the association of different virtual IDS, it can also associate the virtual identity with the real identity, and expand the association of other information. In the follow-up work, we will further study the verification and analysis platform of Internet information. It can intelligently search and analyze the search target, and visually display the search results to realize the combination of virtual and real identities.

ACKNOWLEDGMENT

This work was supported by Ministry of education's humanities and Social Sciences youth fund project supports "Research on social media communication governance based on trust and content risk perception (20 YJC 860032); Education department of science and technology plan projects, No. GJJ161174 in Jiangxi province.

REFERENCES

[1] Jazaieri H, Logli Allison M, Campos B, et al. Content, structure, and dynamics of personal reputation: The role of trust and status potential within social networks. Group Processes & Intergroup Relations, 2019, vol.22, no.7, pp. 964-983. DOI:10.1177/1368430218806056

[2] Urena R, Chiclana F, Carrasco R A, et al. Leveraging Users' Trust and Reputation in Social Networks. Procedia Computer Science, 2019, vol.162, No.12, pp.955-962. DOI:10.1016/j.procs.2019.12.073.

[3] Roy A, Huh J, Pfeuffer A, et al. Development of Trust Scores in Social Media (TSM) Algorithm and Application to Advertising Practice and Research. Journal of Advertising, 2017, vol.46, No.1, pp.1-14. DOI: 10.1080/00913367.2017.1297272.

[4] Cheng X, Fu S, Vreede G J D. Understanding trust influencing factors in social media communication: A qualitative study. International Journal of Information Management, 2017, vol.37, No.2, pp.25-35. DOI: 10.1016/j.ijinfomgt.2016.11.009.

[5] Wang Gang. A survey of domestic network information behavior based on Virtual Community. library, 2017, No.5, pp.47-53.

[6] Nie Yonghao, Luo Jingyue. Perceived usefulness, trust and willingness of social network users to disclose personal information. Library and information knowledge, 2013, No.5, pp. 89-97.

[7] Zafarani R, Liu H. Connecting users across social media sites: a behavioral-modeling approach. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013, pp. 41-49. DOI: 10.1145/2487575.2487648.

[8] Liu S, Wang S, Zhu F, et al. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In Proceedings of the 2014 ACM SIGMOD international conference on Management of data. 2014, pp.51-62. DOI: 10.1145/2588555.2588559.

[9] Nunes A, Calado P, Martins B. Resolving user identities over social networks through supervised learning and rich similarity features. In Proceedings of the 27th Annual ACM Symposium on Applied Computing. 2012, pp.728-729. DOI: 10.1145/2245276.2245413

[10] Gomaa, Ibrahim A., et al. "Virtual Identity Approaches Evaluation for Anonymous Communication in Cloud Environments." International Journal of Advanced Computer conference & Applications, 2016, pp.367-376. DOI: 10.14569/IJACSA.2016.070251.

[11] Gomaa, Ibrahim, et al. "Performance Evaluation of Virtual Identity Approaches for Anonymous Communication in Distributed Environments." Procedia Computer conference, 2017, pp.710-717. DOI: 10.1016/j.procs.2017.05.382