Heterogeneous Identity Expression and Association Method Based on Attribute Aggregation

Wenye Zhu[1,*], Chengxiang Tan[1], Qian Xu[2] and Ya Xiao[1]

[1]*Department of Computer Science and Technology, Tongji University, Shanghai 200092, China*
[2]*Blockchain Research Institute, China Telecom Bestpay Co., Ltd, Shanghai 200080, China*
E-mail: 1310513@tongji.edu.cn
[*]*Corresponding Author*

Received 18 September 2020; Accepted 26 October 2020;
Publication 26 December 2020

Abstract

Existing identity expression methods are often limited in a single security domain, and this is inadequate to meet the cross-domain access requirements of heterogeneous networks. In view of this problem, we propose an index system for the ubiquitous expression of heterogeneous identities, and introduce the concept pair matching based attribute aggregation method by combining the characteristics of heterogeneous identity alliances. The selection of concept pairs considers the original meaning of attribute characteristics, including the lexical level, i.e., class, ontology, label, description, the structural level, i.e., position, distance between nodes, and the semantic level, i.e., formal concept analysis. As for the attribute aggregation, if multiple attributes from a heterogeneous network contain the same or similar concepts, they are considered the same attribute for the user identity in a heterogeneous network. Relevant domain knowledge or heuristic knowledge will adjust the result of attribute aggregation, and the constraint relationship between

Journal of Web Engineering, Vol. 19_7–8, 1267–1290.
doi: 10.13052/jwe1540-9589.197815

conceptual structures are used to adjust and optimize the attribute aggregation set. Based on the identity attribute index system of the heterogeneous identity alliance, the identity similarity evaluation results based on each attribute are generated. When the comprehensively considered identity similarity evaluation result is higher than the empirical threshold, the heterogeneous identity alliance has different trusts for the same user. The experimental results show that our scheme has a better overall aggregation effect on identity attribute aggregation.

Keywords: Heterogeneous identity alliance, attribute aggregation, network identity management, identity expression, trust management.

1 Introduction

The credible evaluation mechanism for heterogeneous identity alliance users has become one of the key issues in the field of trust metrics and trust management in cyberspace. Identity trust negotiation of entities in traditional inter cyberspace is often based on a single trust root [1]. With the development of globalization, the needs for cross-border and cross-trust domain interaction are increasing, and the interconnection and interoperability of heterogeneous network space represented by 5G and Internet of Things is more frequently [2]. The traditional cross-domain trust negotiation mechanism can no longer meet the identity and trust security requirements nowadays.

There are many researches on identity management technology in academia at present, but in the real network space, identity management systems of different architectures and different application domains are coexisting, massive, heterogeneous and polymorphic [3, 4]. The security challenge of network identity management is becoming more and more serious. How to uniformly manage the multi-domain and multi-modal identity of network entities is a challenging problem. A large number of identity management systems are inter-applications. However, the identity management methods among intra-applications are various. This brings many problems, such as, the system integration is inconvenient, and various identity management platforms are not interoperable; multi-dimensional identity authentication experience is poor, management is difficult; credible evaluation of cross-domain identity management is missing; identity privacy information is easily abused and misused; identity information is easily copied and forgery and other issues, network security incidents caused by various identity management

systems and authentication protocol vulnerabilities have emerged in recent years [5–8].

1.1 Contributions

In this paper, we propose a method for constructing a normalized identity attribute index system based on attribute aggregation model for heterogeneous identity expression.

Existing identity expression methods are often limited to a single security domain, and the need for cross-domain access to heterogeneous networks is not sufficient. In this regard, we propose an index system for the ubiquitous and normalized expression of heterogeneous identity. The core steps include attribute extraction and attribute aggregation. Since there are multiple strategies for the attribute aggregation method, the aggregation effect produced by each strategy is also different. Therefore, we combine the characteristics of heterogeneous identity alliances and introduce an attribute aggregation model based on concept pair matching. The selection of concept pairs considers the original meaning of attribute features and the quantitative indicators, including lexical level: class, ontology, label, description, structure level: position between nodes, distance, and semantic level: formal concept analysis. If multiple attributes from a heterogeneous network contain the same or similar concept pairs, then these attributes are treated as the same attribute for the user identity in the heterogeneous identity federation. Relevant domain knowledge or heuristic knowledge adjusts the attribute aggregation result, and the constraint relationship between the concept structures are used to adjust the optimized attribute aggregation set. The attribute index system for heterogeneous identity alliance is used as the evaluation basis, and the identity similarity evaluation result based on each attribute is generated. When the comprehensive similarity evaluation result is higher than the experience threshold, the heterogeneous identity alliance has different trust to the same user. The heterogeneous identity within the domain implements the association.

The experimental results show that the overall aggregation effect of identity attributes is better, but the effect of some special data processing is not ideal. From the comparison experimental results, the method of this paper has a certain improvement in accuracy and recall rate compared with other attribute aggregation models. The percentage of evaluation of the overall aggregation effect of attributes has increased by about 0.15.

1.2 Related Works

1.2.1 Network Identity Management

Network identity management is the key to realizing the combination of real world and cyberspace, providing trust service and behavior regulation, and is the basis for implementing cyberspace governance. The management of network identity has gained great attention in both the industry and researchers. The United States issued the "National Strategy for Authentic Identity of Cyberspace" [9] and the National Action Plan for Cybersecurity [10], and carried out identity authentication services in the fields of e-commerce, education and medical care in many places. The European Union has implemented projects such as STORK and FutureID [11] to promote the development of a unified European electronic identity management legal framework, technical standards, cross-border network identity authentication and trust services.

The current network identity management technology has been widely used in government, banking, e-commerce, social networking and other fields. Users have built their own independent identity management platforms or systems to provide identity management and trust services for personnel and applications within the organization. A large number of identity management infrastructures are centered on applications and information systems. Their identity management is different, their implementation methods are different, and their structures are loose. They form a "island of identity management" for cross-domain identity sharing and business integration. System integration brings a lot of inconvenience. The current network identity management has the following problems: (1) the identity management platform is difficult to communicate with each other; (2) the multi-dimensional identity authentication service experience is poor; (3) the credible evaluation of cross-domain identity management is lacking; (4) The rights management in the heterogeneous environment is complicated; (5) the virtuality of the network entity identity leads to regulatory difficulties; (6) the identity privacy information is easily abused and misused; (7) the identity information is easily copied and forged; (8) the heterogeneous environment identity privacy Data sharing is difficult; (9) Comprehensive analysis of multi-state cross-domain network entity behavior is difficult [12–17]. With the continuous development of network technology, heterogeneous networks with different architectures and different application domains coexist in the network space. How to uniformly manage the multi-domain and multi-modal identity of network entities is a challenging problem.

Building a trust alliance to achieve network identity management has become the consensus of academia and industry. Microsoft's Passport [18], Google's Open Account-based Google Accounts and CICP services [19] have formed a commercial identity management alliance that provides users with consistent identity and account management within the federation, extending the effective domain of user identities from a single organization to the federation. In the needs of new application requirements such as cross-trust alliances, it is necessary to develop heterogeneous identity alliance technologies. Heterogeneous identity alliance is composed of multiple identity management platforms across architectures and application domains. Compared with other identity alliances, it can provide unified, secure and credible, life-cycle identity management and services, which is to improve network space supervision and Effective ways to control and protect the privacy of network identity are the strategic cornerstones for promoting the sound development of the national network economy and safeguarding national network security.

1.2.2 Heterogeneous Identity Alliance

At the end of the 20th century, Microsoft launched the Passport project [18]. By building the Passport.com website, it stores the identity of all users, provides unified authentication services for all authorized sites of Microsoft, and supports users to single sign-on. However, there are serious security risks. Unlike the Passport, the Identity Management Specification [20] issued by the "Freedom Alliance" organization established in 2001 solves the problem of centralized data storage. By establishing a network of multiple identity management systems, users can keep personal information and achieve cross-system Unified identity authentication service, but the alliance is simple to build, relying on the identity provider's service, and the reliability is not high. In addition, IBM's Tivoli Access Manager [21] solution has been enhanced in security, and Novell's iChain technology based on cached reverse proxy [22] has been useful in simplifying management processes and improving service performance. In terms of open source projects, the Oauth Authorization Framework [23] works by delegating user authentication to managed users and authorizing client access to users, which provides authorization processes for web and desktop applications and mobile applications, enabling third-party applications. The program or client gains limited access to user information on the HTTP service (eg, Google, GitHub). OpenID [23] is an open online identity authentication system that allows users to log in to multiple websites using their existing accounts without having to create new

passwords. Instead, they only need to register in advance on an OpenID identity provider's website to enjoy the resources on the OpenID-enabled website. The OpenID system can implement authorization based on the OAuth protocol. With the maturity of technology, ISO, ITU-T, ETSI and other international organizations have successively formulated identity management standards such as SAML, OAuth, OpenID, FIDO, and EU eID unified management framework [24–26].

1.2.3 Attribute Aggregation

The user identity in the traditional single trust domain has a uniform standard attribute set. The identity expression method is the filling of the attribute set. The intra-domain identity evaluation depends on a series of information contained in the identity attribute set [27]. With the increasing demand for cross-domain identity credibility evaluation, how to aggregate identity attributes into unified expression standards has become an infrastructure to solve the above problems.

As a method of knowledge sharing and interoperability between different attributes, attribute aggregation technology has received more and more attention from the academic community. The original intention of attribute aggregation is for information sharing, but it is very difficult to establish a globally common attribute collection. The existing attribute sets are often constructed by different attribute development teams. The attribute construction standards are inconsistent, resulting in attribute heterogeneity [28]. Due to the emergence of attribute heterogeneity, in order to complete the task of information exchange, it is necessary to build a bridge of semantic mapping between attributes. The meaning of attribute aggregation mainly has two aspects: narrow and broad. In a narrow sense, attribute aggregation only includes attribute matching, and only establishes the correspondence between source attributes and target attributes [29]. In a broad sense, attribute aggregation also includes attribute alignment [30] and attribute merging [31]. The broad sense not only establishes the correspondence of the mapping, but also adjusts the structure of the attributes, semantics and axioms according to the correspondence of the mapping, thus forming a new set of attributes [32].

1.3 Paper Organization

The remainder of this paper is organized as follows. In Section 2, we give the ubiquitous representation model of heterogeneous network space user identity. The details of attribute aggregation and heterogeneous identity trust

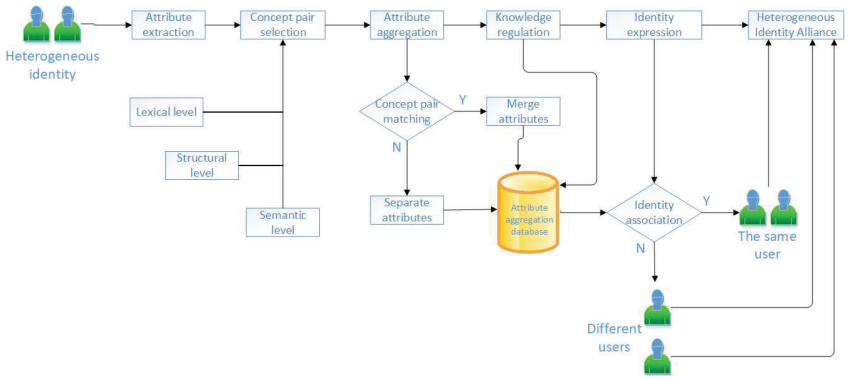

Figure 1 Heterogeneous identity expression flowchart.

association are elaborated in Sections 3 and 4, respectively. Section 5 is dedicated to analyzing the scheme by experiments. Finally, we conclude this paper in Section 6.

2 Model

The ubiquitous representation model for user identity in heterogeneous network space proposed in this paper is shown in Figure 1. The general processes are as follows:

1. Import the cross-domain identity information of users in a heterogeneous network space.
2. Extract the attribute features such as the concept name, semantic name, and concept instance for each user identity.
3. Select the concept pair for attribute aggregation.
4. Aggregate the attribute information of heterogeneous identity based on the concept pair matching algorithm.
5. Adjust the attribute aggregation result by using relevant domain knowledge or heuristic knowledge.
6. Construct a heterogeneous identity ubiquitous normalized expression system based on attribute aggregation.
7. Associate the cross-domain identity of the same user in the heterogeneous identity alliance.

The proposed model collect the identity attributes from different trust domains of heterogeneous networks. The specific identity includes social network accounts, telecommunication network users, electronic ID cards,

etc. The extracted attributes cover the concept names, semantic names and concept instances of user identity.

An attribute feature consists of a number of independent, indivisible primitive meanings, which can be referred to as concept pairs. The selection of concept pairs considers the original meaning set of attribute features, and select attribute aggregation conceptpairfor quantitative indicators by lexical level (class, ontology, label, description), structure level (position between nodes, distance) and semantic level (formal concept analysis).

We use the concept of matching algorithm to aggregate the attribute information of heterogeneous identity. If multiple attributes from a heterogeneous network contain the same or similar concept pairs, then these attributes are treated as the same attribute of the user identity in the heterogeneous identity federation. In contrast, if multiple attributes from a heterogeneous network contain opposite or deviating concepts, then such attributes are considered to be the basic constituent attributes of the user identity in the heterogeneous identity federation.

We use relevant domain knowledge or heuristic knowledge to adjust the attribute aggregation result. The specific method is to verify the identity attribute of the aggregate through the prior knowledge of the specific domain such as telecommunication network, e-government network, social network, etc. The constraint relationships are used to adjust the optimized attribute aggregation set. Based on the concept of the attribute aggregation set formed after the matching, the identity attribute index system of the user's ubiquitous normalized expression in the heterogeneous identity alliance is constructed.

In order to correlate the different identities of the same user in the heterogeneous network space, based on the proposed heterogeneous identity alliance identity attribute index system, we traverse all the attributes contained in the identity to be associated, and use the identity similarity calculation formula to generate the identity similarity based on each attribute. As a result of the identity similarity evaluation, when the comprehensive evaluation of the identity similarity is higher than the empirical threshold, the heterogeneous identity alliance associates the heterogeneous identity of the same user in different trust domains.

3 Attribute Aggregation

3.1 Identity Attribute Concept Selection

In this paper, we use traditional identity attribute extraction technology to collect attribute information from various networks, such as social network

accounts, telecommunication network users, electronic ID cards, et al. The extracted attribute information consists of concept names, semantic names and concept instances of user identity. An attribute feature consists of a number of independent, indivisible primitive meanings, which may be referred to as concept pairs. The selection of concept pairs considers the original meaning of attribute features, and the quantitative indicators used to select attribute aggregation concept pairs including lexical level (class, ontology, label, description), structure level (position between nodes, distance) and semantic level (formal concept analysis).

The concept of vocabulary level can provide strong persuasiveness and judgment for attribute aggregation, and can also provide guarantee for the practicality of attribute aggregation set. Therefore, the vocabulary level diversity index is the basic evaluation index during the process of measuring attribute aggregation. We combine the basic process of identity attribute construction to consider the composition of aggregate indicators at the lexical level. Therefore, the specific indicator content of the vocabulary level diversity includes the definition of the class, the attribute definition, instance, label description and other constraints. The more factors are considered, the more accurate the attribute aggregation at the vocabulary level.

The structure level concept pair selection mainly describes the dependence part of the attribute structure when evaluating the similarity between the source attribute and the target attribute. The dependence part indicate the features of the attribute structure that are used to discriminate the similarity. Similar to the vocabulary level, the aggregation indicators at the structural level are also established from the basic to the comprehensive. Therefore, in the attribute aggregation system, the in-depth evaluation indicators of the structure level mainly include the relationship between the father and the child, the brother relationship, the neighbor relationship, the relationship between the whole and the part, and the distance relationship between the nodes. With the increase of the aggregation index at the structural level, the accuracy of the attribute aggregation at the structural level will gradually increase.

The concept pair of semantic level can accurately describe the semantic relationship between attributes, because it is a progressive and complementary level of vocabulary and structure. The attribute aggregation from different perspectives of the semantic level affects the degree of semantic association between the two attributes and also determines the accuracy of the final attribute aggregation. The main theoretical methods applied in this paper include string similarity, linguistic similarity, formal concept analysis,

information flow and description logic. The above theoretical methods have different semantic strengths. Generally speaking, the higher the semantic strength of a theoretical method, the deeper the semantic degree when mining the attributes, and the higher the accuracy of attribute aggregation under the same conditions.

3.2 Attribute Aggregation Based on Concept Pair Matching

We use the concept pair matching algorithm to aggregate the attribute information of heterogeneous identity. The algorithm is as follows: For the two identity attributes a_1 and a_2, which respectively include n and m concept pairs, namely c_{11}, c_{12}, …, c_{1m} and c_{21}, c_2, …, c_{2m}, the similarity between the two identity attributes a_1 and a_2 can be described as:

$$\mathrm{Sim}(a_1, a_2) = \max_{i=1,2\ldots n, j=1,2\ldots m} \mathrm{Sim}(c_{1i}, c_{2j}). \tag{1}$$

According to the relationship between the concept pairs, a tree-like hierarchy is used to describe the concept pairs. Suppose the distance between two concepts c_1 and c_2 in the tree is denoted as d, and the semantic distance calculation formula is:

$$d = \mathrm{Sim}(c_1, c_2) = \frac{k}{d + k}, \tag{2}$$

where k represents an adjustable parameter.

If multiple attributes from a heterogeneous network contain the same or similar concept pairs, these attributes are treated as the same attribute of the user identity in the heterogeneous identity federation. Conversely, if multiple attributes from a heterogeneous network contain opposite or deviating concepts, such attributes are considered as the basic constituent attributes of the user identity in the heterogeneous identity federation.

3.3 Constructing Normalized Indicators Through Knowledge Adjustment

We use relevant domain knowledge or heuristic knowledge to adjust the attribute aggregation result. The specific method is to verify the aggregated identity attribute through the prior knowledge of the specific domain such as telecommunication network, e-government network, social network, et al. The constraint relationships among the concept structure are used to adjust the optimized attribute aggregation set. Based on the concept of the attribute

aggregation set formed after the matching, the identity attribute index system of the user's ubiquitous normalized expression in the heterogeneous identity alliance is constructed.

3.4 General Identity Expression Based on Character Information

We firstly use the previous proposed identity attribute extraction technology to achieve the import of heterogeneous network space user identity, and then the imported identity information is used to select concept pairs and aggregate attributes. Then the relevant domain knowledge or heuristic knowledge pairs of various heterogeneous networks are applied to adjust the attribute aggregation results. Finally, the following heterogeneous network identity attribute normalization expression method based on the character information is formed.

The virtual heterogeneous network space constructed in this paper contains user identity registration information. The information mainly includes related information of the real name system registration of the character, including mobile phone number, bank card, vehicle, account, personal and family information.

The heterogeneous cyberspace with user virtual network information constructed in this paper mainly refers to the information about the characters registered on the social network, such as QQ, WeChat, Weibo, Zhihu, Facebook, Twitter, Instagram and various types of mailboxes.

The virtual heterogeneous network space with user terminal information constructed in this paper mainly refers to the electronic device related information, such as mobile phones and notebook computers with different operating systems.

The virtual heterogeneous cyberspace with user communication relationship constructed in this paper mainly refers to the communication and interactive relationship, such as group, communication, friend, follow relationships, of network characters in telecommunication network, internet and social network.

The basic information of the virtual heterogeneous cyberspace constructed in this paper mainly refers to the users' basic information and geographical location information. The website covers different aspects, including clothing related (Taobao, Jingdong, Amazon), food related (Mei Tuan, hungry), live related (Ctrip, Yilong) and travel related (Ctrip, 12306, where to go).

The virtual heterogeneous cyberspace with order information constructed in this paper mainly include website order, payment information and logistics

information, such as 12306 train ticket orders, where to go ticket orders, payment orders and logistics orders.

The virtual heterogeneous cyberspace with user search history information constructed in this paper mainly includes Baidu, Google, and Bing corresponding account search.

4 Heterogeneous Identity Trust Association

4.1 Discover Similar Identities

In order to correlate the different identities of the same user in the heterogeneous cyberspace, the general identity expression inheterogeneous identity alliance are used as the evaluation basis. The importances from the same attribute i to the identities a and b in different security domains are denoted as $d_{a,i}$ and $d_{b,i}$, the number of attributes owned by identity a and b is $|I_{a,b}|$, and the number of attributes owned by each of identity a and b is denoted as $|I_a|$ and $|I_b|$ respectively. The important similarity for the same attribute pair with different identity a and b is calculated by Equation (3). The common attribute number based similarity between the identity a and b is obtained by Equation (4).

$$IPT_{a,b} = \frac{\sum_{i=1}^{|I_{a,b}|} (d_{a,i} - \overline{d_a}) \times (d_{b,i} - \overline{d_b})}{\sqrt{\sum_{i=1}^{|I_{a,b}|} (d_{a,i} - \overline{d_a})^2} \times \sqrt{\sum_{i=1}^{|I_{a,b}|} (d_{b,i} - \overline{d_b})^2}}, \tag{3}$$

$$AMT_{a,b} = \frac{|I_{a,b}|}{|I_a| + |I_b| - |I_{a,b}|}, \tag{4}$$

where $\overline{d_a}$ and $\overline{d_b}$ represent the average degree of the importance from all attributes to identities a and b respectively. The similarity calculation formula for identity a and b is obtained as follows:

$$\mathrm{Sim}_{a,b} = \mathrm{IPT}_{a,b} \times \mathrm{AMT}_{a,b}. \tag{5}$$

After traversing all the attributes contained in identities a and b, an identity similarity evaluation result based on each attribute is generated. When the comprehensively evaluated identity similarity evaluation result is higher than the empirical threshold, the two identities tend to be the same. The heterogeneous identity alliance implement association for the same user in different trust domains.

4.2 Confirm Trusted Identity

The trust relationship between user identities is directly related to the accuracy of the association of different identities of the same user, and the trust value between identities is determined according to the similarity of the importance of the attributes to the identities. The attribute importance is similar for identity b and identity a can be generally described as: the deviation of the importance of identity a and identity b for all common attributes is less than the trust threshold E.

$$|d_{a,i} - d_{b,i}| \times \mathrm{AMT}_{a,b} < E. \tag{6}$$

According to Equation (4), the trust relationship between identities can be obtained initially. The formula for calculating the trust value between identity a and identity b is described as follows.

$$\mathrm{Trust1}_{a\to b} = 1 - \frac{\sum_{i=1}^{|I_{a,b}|} \sqrt{(d_{a,i} - \overline{d_a})^2 + (d_{b,i} - \overline{d_b})^2}}{|I_{a,b}| \times \sum_{i=1}^{|I_{a,b}|} [\sqrt{(d_{a,i} - \overline{d_a})^2} + \sqrt{(d_{b,i} - \overline{d_b})^2}]}. \tag{7}$$

Equation (7) does not take into account the importance that a single identity attaches to an attribute, which will result in a very high degree of trust with a very small number of attributes. Therefore, after considering the importance that a single identity attaches to attributes, the formula for calculating the trust value between identities a and b is transformed into:

$$\mathrm{Trust2}_{a\to b} = \begin{cases} E \times \mathrm{AMT}_{a,b} & \mathrm{Trust1}_{a\to b} = 0 \\ \mathrm{Trust1}_{a\to b} \times \mathrm{AMT}_{a,b} & \mathrm{Trust1}_{a\to b} \neq 0 \end{cases}. \tag{8}$$

Equation (8) reflects the direct trust relationship between identities, but there are identities with indirect trust relationships in reality. If identity a trusts identity b and identity b trusts identity c, it can be considered that identity a trusts identity c to some extent. There may be many trust paths for indirect trust. Considering the number of trust paths between identity a and identity c, denoted as $\mathrm{road}(a, c)$, the trust value between identity a and identity c is obtained as follows.

$$\mathrm{Trust3}_{a\to c} = \frac{\sum_{b \epsilon \mathrm{road}(a,c)} \mathrm{Trust2}_{a\to b} \times (\mathrm{Trust2}_{b\to c} \times \beta_d)}{\sum_{b \epsilon \mathrm{road}(a,c)} \mathrm{Trust2}_{a\to b}}, \tag{9}$$

where d is the distance between the two identities, and β_d is the formula for calculating the attenuation index of trust.

$$\beta_d = \frac{\mathrm{DTres} - d + 1}{\mathrm{DTres}}, \quad d \in (2, \mathrm{DTres}). \tag{10}$$

In summary, the final calculation formula for the trust value between identities a and b is:

$$\mathrm{Trust}_{a\to b} = \begin{cases} \mathrm{Trust2}_{a\to b}, & \beta_d = 1 \\ \mathrm{Trust3}_{a\to b}, & 0 < \beta_d < 1 \end{cases} \tag{11}$$

5 Experiment Analysis

5.1 Experimental Data and Scheme

The experimental data used in this article is from the open source data of social networks provided by the snap website of Stanford University's Network Analysis Lab. The data mainly includes Facebook's identity attribute instance, and there is a large amount of node data describing the association of attributes between identities. The data source has 4039 nodes and 88234 edges. The relationship examples involve character relationships and organizational relationships.

The experimental method in this article is compared with the PROMPT mapping method based on string matching [30] and the RIMOM mapping method based on Bayes decision theory [31]. The open source data of social network provided by the snap website of the Stanford University Network Analysis Lab is used as the experimental object [33]. In the dataset, the No.101 identity is treated as a reference attribute concept pair, the artificially annotated attribute aggregation results are used as the evaluation criteria. The accuracy rate and recall rate are obtained by comparing the results of the manual annotation with the effect of the algorithm aggregation, and the F value is used as a comprehensive index to measure the quality of the attribute aggregation result. We extract 41 identity attribute tuples from the Social circles: Facebook dataset. Generally, we use the No.101 identity as the reference attribute concept pair, and the others are deformed attribute sets with one or some features missing. The test set is divided into three categories, namely simple test, system test and real identity test.

1. Simple test (1XX): This group of test set mainly includes the identities 101 to 105, mainly for the concept test. The test is the identities 101 and their related identities. These identities are limited to the reference identities under the normalized index.
2. System test (2XX): This group of test set mainly includes 201 to 250 identities, mainly to test the robustness of the system after the identity or some attributes are lost.

3. Real test (3XX): This group of test set mainly includes 301 to 305 identities. It mainly tests how effective the 5 real identities are in the attribute aggregation application.

$$R = \frac{\text{Number of correct aggregations}}{\text{Number of reference aggregations}} \tag{12}$$

$$P = \frac{\text{Number of correct aggregations}}{\text{Actual total aggregated quantity}} \tag{13}$$

$$F = \frac{2{*}R{*}P}{R+P} \tag{14}$$

5.2 Experiment Platform

The environment of this experiment was built on Lenovo (ThinkServer) RS260 1U rack server. The CPU model is Xeon E3-1220V6, with 8GB of memory, and the operating system is Ubuntu 18.04.3. We use Python and Java for data pre-processing and attribute aggregation model program writing. The API and Jena interface in Java are used to analyze the identity attribute relationship to be aggregated. The Java WordNet Library is used to read the semantic dictionary WordNet component library. Stanford University Protégé tool was developed for the operational editing of identity attributes.

5.3 Experimental Results

The experimental results are shown in Table 1.

This experiment uses three different types of identity attributes for test on the snap data source: 1XX, 2XX, and 3XX. Table 1 shows the specific results of the experiment on the snap data source. In the 1XX data set, the No.103 data cannot obtain the experimental results. There is no associated attribute between identities. In the 2XX data set, the five sets of identity attribute, i.e., 216, 226, 238, 239, and 244, have no aggregation result output, so they are not used as experimental test objects.

Table 2 is the average accuracy rate, average recall rate, and F value of the three methods in the three groups of 1XX, 2XX, and 3XX data, where the F value is a comprehensive measure of the quality of the attribute aggregation results based on the accuracy rate and recall rate.

The PROMPT method only uses string matching and semantic dictionary combination to calculate the similarity relationship between identities. The RIMOM method, based on the connected relationship between attribute

Table 1 Experimental results on snap data sources

System	PROMPT		RIMOM		Attribute Aggregation	
Test	Prec	Rec	Prec	Rec	Prec	Rec
101	0.97	1.00	1.00	1.00	1.00	1.00
102	0.91	1.00	1.00	1.00	0.93	1.00
103	NAN	NAN	0.00	0.00	NAN	NAN
104	0.98	1.00	0.00	0.00	1.00	1.00
105	0.93	1.00	1.00	1.00	0.98	1.00
201	1.00	0.06	1.00	1.00	1.00	1.00
202	1.00	0.06	0.91	0.67	1.00	0.86
203	0.96	1.00	1.00	1.00	1.00	1.00
204	0.94	1.00	1.00	1.00	1.00	1.00
205	0.73	1.00	1.00	0.98	1.00	0.97
206	0.87	0.84	1.00	0.98	0.96	0.96
207	0.81	0.85	1.00	0.98	0.89	0.89
208	0.98	0.95	0.99	0.98	1.00	0.99
209	0.92	1.00	0.97	0.74	1.00	0.99
210	0.94	1.00	0.96	0.87	1.00	0.97
211	0.75	1.00	1.00	1.00	1.00	0.99
212	0.87	0.84	1.00	1.00	1.00	0.98
213	0.94	0.92	1.00	1.00	1.00	0.99
214	0.96	1.00	1.00	1.00	0.96	0.92
215	1.00	0.06	1.00	1.00	0.91	1.00
216	0.87	0.84	0.92	0.92	NAN	NAN
217	0.95	0.90	0.92	0.92	0.94	1.00
218	0.95	0.92	1.00	1.00	0.89	1.00
219	0.83	0.24	0.96	0.96	1.00	1.00
220	0.81	0.29	0.91	0.92	0.93	1.00
221	1.00	0.21	0.78	0.92	1.00	1.00
222	1.00	0.21	0.92	0.92	1.00	1.00
223	0.96	1.00	0.91	0.92	1.00	0.78
224	0.81	0.85	0.72	0.93	1.00	0.69

(*Continued*)

Table 1 Continued

System	PROMPT		RIMOM		Attribute Aggregation	
Test	Prec	Rec	Prec	Rec	Prec	Rec
225	1.00	0.06	0.92	0.52	1.00	1.00
226	NAN	0.00	0.81	0.68	0.92	1.00
227	0.98	1.00	0.94	0.41	0.93	1.00
228	0.87	1.00	0.81	0.55	1.00	0.74
229	1.00	0.21	0.90	0.51	1.00	0.45
230	0.81	0.85	0.91	0.59	0.83	0.79
231	0.78	1.00	1.00	0.27	0.83	0.65
232	0.72	1.00	0.94	0.45	1.00	0.98
233	0.75	1.00	0.82	0.56	1.00	1.00
234	0.83	1.00	0.91	0.59	1.00	0.98
235	0.74	1.00	0.86	0.45	1.00	0.98
236	1.00	0.02	0.92	0.35	0.87	0.71
237	1.00	0.02	1.00	0.26	0.82	0.56
238	NAN	0.00	0.83	0.45	1.00	0.98
239	NAN	0.00	0.91	0.34	0.90	0.43
240	0.92	1.00	0.92	0.92	0.85	0.25
241	0.99	0.91	0.94	0.45	1.00	0.27
242	0.86	0.96	1.00	0.27	0.87	0.47
243	1.00	0.02	0.96	0.96	0.89	0.29
244	NAN	0.00	0.96	0.87	1.00	0.50
245	0.00	0.00	0.90	0.51	0.79	0.59
246	0.00	0.00	0.92	0.35	0.87	0.78
247	1.00	0.02	0.81	0.55	0.96	0.51
248	1.00	0.02	1.00	1.00	1.00	0.27
249	1.00	0.02	0.91	0.59	0.93	0.59
250	0.00	0.00	0.90	0.51	0.89	0.57
301	0.91	0.57	0.97	0.65	0.97	0.67
302	0.94	0.46	0.81	0.42	0.86	0.55
303	0.72	0.42	0.63	0.68	0.89	0.75
304	0.91	0.67	0.91	0.96	0.91	0.94
305	0.68	0.49	0.69	0.49	0.99	0.95

Table 2 Comparison of different algorithms on the snap data source

System	PROMPT		RIMOM		Attribute Aggregation	
Test	Prec	Rec	Prec	Rec	Prec	Rec
1XX	0.95	1.00	0.6	0.6	0.98	1.00
2XX	0.85	0.56	0.93	0.73	0.95	0.80
3XX	0.83	0.52	0.80	0.64	0.92	0.77
AVG	0.877	0.693	0.777	0.657	0.95	0.857
F	0.774		0.712		0.901	

nodes and edges, transforms the attribute elements into nodes on the graph. Most of the included matchers use string matching technology, and the rest use synonym dictionary technology. However, the proposed method in this paper uses a combination of semantic dictionary and statistical technology when matching concept pairs, which can effectively solve the problems that caused by the above semantic dictionary. We select identity concept pairs before attribute aggregation. Not only can we find similar relationships between similar elements, such as concepts to concepts, attributes to attributes, but also concepts to attributes, instances to concepts, attributes to attribute values, the affiliation between non-homogeneous elements, et al. The proposed method also allows the existing of some isolated identities in heterogeneous identity alliances, which is of great help in finding all possible attribute aggregation sets.

From the experimental results, the overall aggregation effect of identity attributes is better in this paper, but some special data processing effects are not ideal. The reason is that similar attributes are extremely scarce. From the comparative experimental results in Table 2, the method in this paper has improved the accuracy and recall rate compared to the attribute aggregation based on PROMPT and RIMOM mapping. From the perspective of the overall aggregation effect of the attribute, that is, the attribute aggregation result that measured by F value in the experimental scheme in this paper increased by 0.127 and 0.189 respectively. However, due to the incomplete processing of special data with extremely rare attributes of the same type in this experiment, it has a certain impact on the recall and accuracy.

6 Conclusions

User identity attribute classification and trust structure of heterogeneous networks are various. Attribute aggregation technology is the link between

different identities. It is the uniform expression of user identity attribute information across trust domains in the same comparable evaluation dimension.The main purpose of attribute aggregation is to realize knowledge exchange between different identities, and to achieve information reuse and interoperability.The essential problem is the algorithm of similarity, i.e., how to determine the correspondence between two attributes.We use the method of constructing attribute concept pairs as the basis for attribute aggregation. The entry points mainly include lexical level (class, ontology, label, description, etc.), structural level (position, distance between nodes, etc.) and semantic level (formal concept analysis, etc.), the accuracy of concept selection will directly affect the accuracy of attribute aggregation.In addition, in the construction of a ubiquitous and normalized expression index system for user identity in heterogeneous identity alliances, the concept of attribute aggregation process continuously modifies the matching results (based on prior knowledge and heuristic algorithm adjustment) to increase the conviction of matching results.

Compared with the attribute aggregation based on PROMPT and RIMOM mapping, the attribute aggregation experiment designed in this paper has improved the accuracy and recall rate, and the overall aggregation effect of identity attributes is better. In summary, the ubiquitous expression of user identities in heterogeneous identity alliances includes the introduction of heterogeneous network user identities, attribute feature extraction, concept pair selection, attribute aggregation based on concept pair matching, knowledge adjustment of attribute aggregation results, and heterogeneous identity attribute system construction and heterogeneous identity associations based on identity similarity calculation.

Acknowledgments

This work was supported by the National Key Research and Development Program of China under Grants 2017YFB0802300 and 2017YFC0803700.

References

[1] E. Sanzi, S. A. Demurjia and J. Billings, 'Integrating Trust Profiles, Trust Negotiation, and Attribute Based Access Control,' in 2017 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud), San Francisco, CA, USA, 2017, pp. 177–184.

[2] I. Lee and K. Lee, 'The Internet of Things (IoT): Applications, investments, and challenges for enterprises,', Business Horizons, vol. 58, no. 4, pp. 431–440, 2015.
[3] D. C. Hardt, 'Auditable privacy policies in a distributed hierarchical identity management system,' U.S. Patent 9,245,266, Jan. 26, 2016.
[4] G. J. Ahn, 'Identity selector for use with a user-portable device and method of use in a user-centric identity management system,' U.S. Patent 9,935,935, Apr. 3, 2018.
[5] S. M. Smith and D. Khovratovich, 'Identity System Essentials,'. Evemyrn, 2016, pp. 16.
[6] K. Fragkiadaki, S. Levine, P. Felsen, et al. 'Recurrent network models for human dynamics,' in Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 2015, pp. 4346–4354.
[7] M. Kohtamäki, S. Thorgren and J. Wincent, 'Organizational identity and behaviors in strategic networks,' Journal of Business & Industrial Marketing, vol. 31, no. 1, pp. 36–46, 2016.
[8] M. Giroux, 'From Identity to Alliance: Challenging Métis Inauthenticity through Alliance Studies,' Yearbook for Traditional Music, vol. 50, pp. 91–118, 2018.
[9] J. Werner, C. M. Westphall, C. B. Westphall, 'Cloud identity management: A survey on privacy strategies,' Computer Networks, vol. 122, pp. 29–42, 2017.
[10] F. Gonçalves, B. Ribeiro, O. Gama, et al. 'Hybrid model for secure communications and identity management in vehicular ad hoc networks,' in 2017 9th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Munich, Germany, 2017, pp. 414–422.
[11] P. Modesti, T. Gross, S. Mödersheim, et al. 'Security Evaluation of FutureID,' FurtureID, 2015. [Online]. Available: http://www.futureid.eu/data/deliverables/year3/Public/FutureID_D12.03_WP12_v1.0_Security_Evaluation.pdf.
[12] A. K. Pathan, 'Security of self-organizing networks: MANET, WSN, WMN, VANET,' CRC press, 2016.
[13] J. W. Rittinghous and J. F. Ransome, 'Cloud computing: implementation, management, and security,' CRC press, 2017.
[14] P. Gao, J. S. Baras and J. Golbeck, 'Semiring-based trust evaluation for information fusion in social network services,' in 18th international

conference on information fusion (Fusion), Washington, DC, USA, July, 2015, pp. 590–596.
[15] Y. Du, X. Du and L. Huang, 'Improve the collaborative filtering recommender system performance by trust network construction,' Chinese Journal of Electronics, vol. 25, no. 3, pp. 418–423, 2016.
[16] H. Han, A. K. Jain, F. Wang, et al. 'Heterogeneous face attribute estimation: A deep multi-task learning approach,' IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 11, pp. 2597–2609, 2017.
[17] B. Eze, C. Kuziemsky C and L. Peyton L, 'A patient identity matching service for cloud-based performance management of community healthcare,' Procedia computer science, vol. 112, pp. 287–294, 2017.
[18] R. Oppliger, 'Microsoft .net passport: A security analysis,' Computer, vol. 36, no. 7, pp. 29–35, 2003.
[19] B. Ellin, "About openlD," [Online]. Available: http://www.openidenabled.com/openid/about-openid, 2006.
[20] L. Alliance, 'Liberty alliance project,' [Online]. Available: http://www.projectliberty.org 24 (2002).
[21] K. Günter, 'Access control with IBM Tivoli access manager,' Acm Transactions on Information & System Security, vol. 6, no. 2, pp. 232–257, 2003.
[22] Novell, 'Identity cloud,' [Online]. Available: http://novell.com/ichain.
[23] D. Hardt, 'The OAuth 2.0 authorization framework,' [Online]. Available: http://tools.ietf.org/html/rfc6749.
[24] T. Martens, 'Electronic identity management in Estonia between market and state governance,' Identity in the Information Society, vol. 3, no. 1, pp. 213–233, 2010.
[25] H. Kubicek, 'Introduction: conceptual framework and research design for a comparative analysis of national eID Management Systems in selected European countries,' Identity in the Information Society, vol. 3, no. 1, pp. 5–26, 2010.
[26] T. Lenz and B. Zwattendorfer, 'A Modular and Flexible Identity Management Architecture for National eID Solutions,' International Conference on Web Information Systems and Technologies. 2015, pp. 321–331.
[27] A. Barbir, 'Identity attribute exchange and validation broker,' U.S. Patent 8,935,808, Jan. 13, 2015.
[28] J. Wang J, X. Zhu, S. Gong, et al. 'Transferable joint attribute-identity deep learning for unsupervised person re-identification,' Proceedings of

the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA, 2018, pp. 2275–2284.
[29] G. Bella, F. Giunchiglia, F. McNeill, 'Language and domain aware lightweight ontology matching,' Journal of Web Semantics, vol. 43, 43, pp. 1–17, 2017.
[30] L. Asprino, V. Presutti, A. Gangemi, et al, 'Frame-based ontology alignment,' Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 2017.
[31] M. Fahad, 'Merging of axiomatic definitions of concepts in the complex OWL ontologies,' Artificial Intelligence Review, vol. 47, no. 2, pp. 181–215, 2017.
[32] Z. Yan, L. Zhang, W. Ding W, et al, 'Heterogeneous data storage management with deduplication in cloud computing,' IEEE Transactions on Big Data, vol. 5, no. 3, pp. 393–407, 2017.
[33] J. Leskovec, J. Mcauley, 'Learning to discover social circles in ego networks,' Advances in neural information processing systems, 2012, Tillamook, USA, pp. 539–547.

Biographies

Wenye Zhu received his B.S. degree from the Department of Computer Science and Technology, Tongji University, China, in 2013. He is now pursuing the Ph.D. degree at Department of Computer Science and Technology, Tongji University. His research interest is information security.

Chengxiang Tan received the Ph.D. degree in engineering from Northwestern Polytechnic University, China, in 1994. He is currently a Professor of computer science with Tongji University. His research interests include cyber security, privacy preservation and data analyzing.

Qian Xu received the Ph.D. degree from the Department of Computer Science and Technology, Tongji University, China, in 2020. He is now working in Blockchain Research Institute, China Telecom Bestpay Co., Ltd. His research interests include cryptography and cloud security.

Ya Xiao received the B.S. degree in School of Computer Science and Engineering, Tongji University, Shanghai, China, in 2015. She is currently pursuing her Ph.D. degree in Tongji University of Computer Science and Engineering, Shanghai, China. Her research interests include social networking and natural language processing.