Research on Two-Way Human Resource Recommendation Method Considering User Privacy

Jiangjing Lin[✉], Chunliang Fu, Linhua Gong, and Ming Guo

Wuhan Second Ship Design Institute, Wuhan 430205, Hubei, China
`linjiangjing@whhwtech.com`

Abstract. Traditional two-way human resource recommendation method has the problem of incomplete data anonymous process, which leads to low recommendation accuracy. Therefore, a two-way human resource recommendation method considering user privacy is designed. Collect recruitment information from human resources big data, and convert file formats in batches. We vectorizing each piece of information, build user privacy protection model, calculate entropy redundancy of data attribute value, optimize data anonymous process, calculate similarity between users according to preference score, and set two-way recommendation mode of human resources. Experimental results: The average recommendation accuracy of the two-way human resource recommendation method in this paper and the other two two-way human resource recommendation methods are 55.645%, 48.363% and 48.267% respectively, indicating that the two-way human resource recommendation method has a wider application scope on the basis of combining user privacy.

Keywords: user privacy · human resources · two-way recommendation · enterprise recruitment · employment information · data anonymity · big data

1 Introduction

With the emergence of new technologies such as mobile Internet, cloud computing and the Internet of Things, human society has entered the era of big data. With the development of the Internet, especially the continuous expansion of the number of users of social networks, there are hundreds of thousands of new users and millions of page views every minute. These huge data, in a sense, are a very valuable resource. China has a large population base, and big data is valued by all walks of life in China and is widely used in China. The future development will be more rapid, and it is imperative for enterprises to use big data. Online recruitment refers to a series of recruitment activities in which enterprises use the Internet to publish recruitment advertisements, job seekers apply for jobs online, recruiters screen resumes online, and send online assessments through the Internet. Online recruitment has the advantages of strong timeliness, wide coverage, low cost, and simplified process. In recent years, the flow of talents in my country has accelerated, and the number of Internet users has continued to increase, which

has promoted the further development of online recruitment. With the wide application of online recruitment, the number of users of comprehensive recruitment websites such as 50Qianqian.com, Zhaopin.com and Liepin.com is also increasing rapidly and rapidly. Exponential growth requires recruiters to spend a lot of time screening job applicants' resumes and responding to job seekers' information, which places higher requirements on the data processing capabilities of recruiting companies. Although traditional online recruitment has incomparable advantages over other recruitment channels, it is incapable of handling massive data, and the efficiency and accuracy of data processing need to be improved, which cannot meet the talent recruitment needs of recruiting companies.

With the acceleration of THE network by 5G technology, the further development of artificial intelligence technology and cloud computing technology has greatly promoted the early arrival of the era of big data. While Internet big data brings many conveniences to our lives, it also invisibly exposes personal data privacy on servers without security protection and even directly publishes it on the Internet in plaintext. For example, many websites require applicants to manually fill out and upload their resumes in accordance with the resume format of the website. Filling out resumes in different formats on different platforms brings great burden to job seekers, resulting in a waste of time and resources. Examples like this show that these recommendations are based on the user's personal history data or other behavioral data analysis and algorithms. Therefore, in the information age, while enjoying the convenience brought by all kinds of intelligent software, the problem of personal data privacy disclosure is fundamentally unavoidable. However, users' personal privacy can be protected to a certain extent through privacy protection technology [15] [9] [3]. Moreover, in the follow-up recommendation, the website mostly makes matching recommendation according to the desired position and professional skills filled in by job seekers. The electronic resume uploaded by job seekers is rarely analyzed and utilized, and the accuracy of recommendation is average. Usually, the recommendation results have to be identified and marked by manpower consumption to have a certain degree of suitability. Therefore, a more accurate, safe and objective two-way recommendation method of human resources is urgently needed to solve the above outstanding problems.

2 Research on Two-Way Human Resource Recommendation Method Considering User Privacy

2.1 Collecting Recruitment and Employment Information

Through the service terminal machine, the personal information of job seekers is collected, including the applicant's name, gender, educational background, work experience, job intention and other information. Since most companies do not have uniform style and format requirements for resumes, job seekers may choose different templates and formats based on personal preference. The format of resumes may also be tables, images, text or other content. Then collect the basic information of the post from the employer, including the job content, working years, gender requirements, age requirements, monthly salary, welfare and other information of the post. In order to facilitate the subsequent text analysis, the electronic resume is converted in batches through the document batch conversion tool, and the files including DOC, DOCX, PDF and other formats

are converted into text files for storage. Part of the simple text in the structured text is used for preliminary data screening, such as job nature, job title education background, gender, and part is used for numerical matching, such as salary, working location, etc. Then, the expression formula of numerical matching after discretization is:

$$sim = \sum_{e=1} sim\left(h_e - \frac{1}{\alpha}\right) \tag{1}$$

In formula (1), h represents structured text, α represents text quantity, and e represents matching coefficient. For unstructured text matching, two text matching methods are adopted. The first method is to calculate text similarity using TF-IDF algorithm. From the perspective of the broad framework, the content of the resume usually contains personal information, preference and intention, work experience, educational background, skills certificate and self-evaluation, etc. These modules are usually arranged side by side on the resume according to the module classification label, but sometimes the content between different modules will also appear crossover. Finally, the weighted sum of the calculated values for structured and unstructured text matches is performed. The formula for calculating text similarity is:

$$sim(P, Q) = \frac{1}{\gamma} \times sim\beta + \gamma \times \beta \tag{2}$$

In formula (2), P, Q represent the resume text of the job seeker and the recruitment text of the enterprise respectively, γ represents the enterprise demand coefficient, and β represents the recruitment data dimension. On this basis, the project has built a human resources employment platform, providing channels for job seekers and employers to find jobs and recruit workers. At present, the platform runs well and has accumulated some information about job seekers' browsing, collecting and applying for positions. The recruitment information of an enterprise usually contains the basic information of the position of the enterprise, the requirements of the enterprise for personnel, the job description of the enterprise, the introduction of the development of the company and other modules. There are overlapping parts in the job description and preference modules, because some enterprises prefer to indicate the job preference in the job requirements. Among them, the historical behavior data of job seekers includes job seeker number, post number, behavior type, behavior occurrence time and other information. Based on the above description, complete the steps of collecting recruitment and employment information.

2.2 Building User Privacy Protection Model

Big data contains huge commercial value. At present, all walks of life are doing big data analysis and mining. They explore potential value in their own databases or data published on the Internet to maximize the utility of data implicit information. Privacy is all the things that an individual does not want others to know about the area of his life [13] [14]. While enjoying the convenience brought by big data mining to production and life, it also inevitably discloses people's privacy. Privacy right refers to the basic

human right that the peace of private life and the secret of private information enjoyed by natural persons are protected according to law, and are not illegally disturbed, known, collected, utilized or publicized by others. How to not only protect users' data privacy, but also keep the original statistical characteristics of data as much as possible has become a research hotspot in the field of data privacy security [10]. Internet privacy is the embodiment of privacy right in cyberspace. It means that people have the right to know, choose and have reasonable access control over personal data, private information and personal field online. However, in real application scenarios, privacy protection and data mining are often contradictory to each other. How to improve the utilization rate of big data without disclosing user privacy can be solved by seeking a balance between the two [12] [7]. It also guarantees their safety and the right to seek judicial assistance, such as prohibiting the disclosure of sensitive information related to individuals on the Internet, and making defamatory remarks. After the data table is not processed or released through certain processing, attackers have the following common attack methods in order to obtain the user's private data or find the person corresponding to the sensitive attributes they are interested in. Identity information usually refers to the data that can represent specific individuals in real life, such as the user's name, age, gender, ID card number, nationality, occupation, etc., which is usually confidential information. Private data refers to information about personal behavior or value orientation released by users on social networks, such as online diaries, albums created, videos and images, which are generally restricted access information. Link attack: it is the most common and basic data privacy attack. It is generally aimed at data tables with simple IDS. On the basis of the obtained data, it is necessary to process these data, and then weight these data by entropy method, so as to reflect the influence degree of this data on the final result more objectively, and then get the comprehensive score. Finally, each piece of information is vectorized, and the database required by the experiment in this paper is finally generated. Taking the application information table as an example, assuming that there are i students in the data table and each student has j attributes, the matrix is as follows:

$$L = \begin{bmatrix} t_{11} & \cdots & t_{1i} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{ij} \end{bmatrix}_{i \times j} \tag{3}$$

In formula (3), t represents the information entropy of attribute value. Information owners can restrict access groups and operations through access management methods and tools provided by the website, such as setting friends lists, authorized information reprinting or copying conventions, etc. The specific process is that other data tables containing the groups to be attacked are first obtained through some channel. Then, according to the comparison analysis and matching of qids, the processed ids are restored. Finally, the desired privacy information is obtained through this link relationship. User-to-user relationship refers to the relationship between users in social networks, such as whether there is a connection and how close it is. Such relationship may exist in the same social service website or in multiple different types of social service websites. On the basis of formula (3), the attributes of the application information are reduced, and

formula (4) is obtained as follows:

$$L' = \frac{\max \phi - \min \phi}{G\{\phi - 1\}} \tag{4}$$

In formula (4), ϕ represents the contribution proportion and G represents the relative value of data attributes. Homogeneity attack is an attack mode evolved from the above link attack, which mainly aims at the attacker to master some known information. The basic principle is that the processed data show that the same equivalent class often has the same or similar sensitive attributes. If a user is known to be in it, the sensitive attributes that need to be protected will be leaked. Social networking sites not only have users' insiders in the social circle, but also many malicious attackers. The property of users' personal information in the network makes more and more malicious attackers attempt to use users' personal information to seek benefits. At the same time, with the development of network technology and information communication technology, it also provides a convenient means for them to infringe on network privacy. Background knowledge attack refers to that the attacker researches the social background of the attacked and obtains certain individual user knowledge background information through comprehensive analysis before the attack. Its basic principle is that when the attacker cannot uniquely determine the user attributes, he can predict the corresponding relationship between the user and the sensitive attributes with a high probability based on the user knowledge background information he has mastered. Based on this, the steps of constructing the user privacy protection model are completed.

2.3   Optimizing the Data Anonymous Process

Simple anonymous Posting only hides the identity of users in social networks without modifying the structure of social network graph, which increases the effectiveness of social network graph structure to some extent. Signcrypto is a low-cost and effective method to realize both authentication and confidentiality in cryptography. Simple anonymous publishing is safe if the attacker has no background knowledge. Ring signature encryption allows members to represent ring signature encryption messages and protect the real identity of the signer. Unlike a group, a ring has no central administrator. However, in practical application scenarios, since most nodes in the graph have unique structural features and the background knowledge of the attacker is very complex, it is easy for the attacker to obtain relevant information about the target node through other channels and implement de-anonymization of the node. Attribute base ring signcryption: a ring member represents the sensitive attribute that the ring uses its own attribute signcryption message to protect its real identity. Other members cannot identify the signcrypter's exact identity, but can confirm it as an intra-ring member. In social networks, users are usually required to register with identity information, such as email address, name, etc. Blind signature is an interactive protocol between the user and the signer, the user signs a message, the signer cannot get the message and signature results. In order to prevent the disclosure of such information, Internet service providers adopt anonymity technology, that is, to hide the relevant information that can identify users in the published data to protect users' privacy, and at the same time, try to retain the graph

structure of the original social network to ensure the validity of the graph information in the social network. Blind signature based on identity realizes a probabilistic polynomial time interactive signature protocol between signer and user. Signer uses the private key to sign the blind message and verify whether the identity of the signature is legitimate. According to the different information possessed by the attacker, the de-anonymising attack is divided into two types: node precise query and subgraph knowledge query. Media events play an important role in promoting information policies of different companies. A company with a strict privacy policy may use it as a publicity tool to attract consumers, while a company with a poor privacy policy will lose customers. The greater the dispersion degree of variation of each data value under the attribute, the greater the weight should be assigned to the data; otherwise, the more invariable the data is, the less contribution it makes. Therefore, information entropy redundancy is defined as the difference between the maximum information and the actual information. The expression formula for entropy redundancy of calculated data attribute values is as follows:

$$F_\sigma = 1 - \frac{v(1-\sigma)}{K} \tag{5}$$

In formula (5), v represents the difference degree of applicant's data information, σ represents contribution amount, and K represents weight coefficient. Node exact query is to find out the unique non-equivalence isomorphic nodes in the graph. Under the condition that the attacker has enough information, they can be uniquely identified so as to de-anonymize. In particular, as privacy concerns increase, users are more likely to provide incomplete information to a site and may directly reduce the likelihood that a user will create a profile or use the site. On the other hand, the query and identification of nodes with isomorphism of equivalence classes need to be judged probabilistically according to the number of nodes with isomorphism of equivalence classes in the graph, so there is some uncertainty in the de-anonymization. Market regulation plays a role in the selection and use of social network websites mainly through media publicity or public evaluation of website privacy Settings and information request policies. This query, on the one hand, requires the attacker to fully grasp the neighbor node information of the de-anonymized target node. In fact, market regulation is to influence users' concerns about their privacy through the dissemination of people around them or those who have a great influence on them. It is an information transmission mechanism, that is, to influence users' concerns about their privacy through subjective norms. On the other hand, if there are many connections between a node and its neighbors, the query will describe a large subgraph, and it is not easy to find the non-equivalence isomorphic nodes. In particular, as privacy concerns increase, users are more likely to provide incomplete information to a site and may directly reduce the likelihood that a user will create a profile or use the site. In contrast, a subgraph query is a global query that considers the presence or absence of a subgraph around a target node. It describes the extent to which an attacker can distinguish and de-anonymize nodes with limited external information. Different subgraphs can be considered as different strategies for attackers to seek target node de-anonymization in social network graphs based on different side facts.
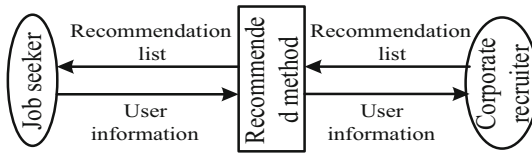
Fig. 1. Schematic diagram of two-way recommendation process

2.4 Setting the Two-Way Hr Recommendation Mode

The two-way recommendation method mainly relies on information retrieval and information filtering, and recommends the recommendation items that are more likely to be accepted to users according to their historical records, which can be divided into user-based recommendation algorithm and object-based recommendation algorithm. The purpose of two-way recommendation is to recommend the employer and the employer respectively [8] [2]. For employment-oriented recommendation, that is, to recommend suitable enterprises to fresh graduates [5]. The user-based recommendation method requires that the user's historical preference score can be obtained, and the similarity between users can be calculated according to the preference score, so as to recommend items that exist in a user's favorite list but not in the user's favorite list to other users with high similarity. First according to the degree of similarity in the previous section to get a list, find out the similarity from high to low respectively every past life, and because in the database is already signed up for the past life and one-to-one, so he can recommend to seniors and the highest similarity of old enterprises, the final result to use Top - N import two-way recommended method, this is the recommended method based on content. Based on the item recommendation method, users can recommend items with high similarity to their favorite items by calculating the relatively static item similarity in the database. For recruiters' recommendation, that is, to recommend suitable students to fresh graduates. Two-way recommendation methods are widely applied in one-way recommendation scenarios such as news recommendation, topic recommendation, scenic spot recommendation, and interest tag recommendation. First of all, the recruiter with previous signing records can be recommended directly. If the newly added recruiter finds a recruiter with the most similar record as his reference value through the similar clustering algorithm described in the previous section. That is to say, in the selection of candidate recommendation objects, the two-way recommendation method should not only make the resources satisfy the user's subjective preference, but also make the user's objective attribute resources satisfy the subjective preference expressed by the user. For enterprises, this two-way resource recommendation method can also meet the needs of enterprise talent selection. Only in this way, the method recommended by human resources can really meet the needs of users. The two-way recommendation process is shown in Fig. 1.

As can be seen from Fig. 1, the recommendation on the recruitment website is the mutual recommendation of the information of the job seeker and the enterprise. Both the job seeker and the enterprise have their preference for the other party, and the whole recruitment process presents a state of mutual selection. However, the two-way recommendation method only considers the unilateral demand of users for items and

cannot meet the demand of two-way recommendation in the field of recruitment. Then, the neighbor set is generated for the former students who have signed the contract with the recruiter, and the similarity of the students to be recommended is calculated respectively. Finally, the results are weighted and sum processed to generate the recommendation list, and the final results are imported into the two-way recommendation method using top-N method, which is an object-based collaborative filtering recommendation algorithm. In order to achieve mutual satisfaction in recommendation, the concept of reciprocal recommendation is proposed and applied in online dating. After the above process, the two-way employment two-way recommendation method is formed, which lays a good foundation for the system integration in the next chapter. By comparison with the two-way recommendation method, it can be found that the two-way recommendation method considers the common attributes of both users and recommended candidates, and the recommendation effect is improved [6] [1]. Six modules are obtained through module label classification, including personal information, preference and intention, educational background, work experience, self-evaluation and other information. Users have a strong psychological protection for personal information and reject the use of historical preference records. However, users' preferences need to be obtained through personal information or historical behavior records, so it is difficult to obtain and analyze information. The basic personal information module usually appears at the beginning of the resume, which can describe the basic personal information of the applicant, including name, gender, age, address, mobile phone number, email, work experience, and seven specific information attributes. The job intention module usually appears in the resume with clear job objectives. In two-way recommendation, both recommancers are willing to provide some basic personal information and preference information to get better recommendation. For example, job seekers will take the initiative to upload their resumes on the recruitment website and fill in their job search preferences and other information [11] [4] Most resumes on recruitment websites have this module, which contains five specific information attributes of job nature, expected work location, expected position, expected salary and job hunting status. The education experience module describes the process and level of the job seekers' education. At the same time, the enterprise will also take the initiative to upload the basic job information and job requirements to the website for job seekers to choose the enterprise position, these available and large amount of data information for the realization of the website personnel and job mutual recommendation function has laid a foundation. Based on this, complete the steps of setting the two-way recommendation mode of human resources.

3   Experimental Test

3.1   Building an Experimental Environment

The development environment of the whole experiment scheme is Windows and the test environment is MacOS(in order to improve the system availability, the two most commonly used computer operating systems are selected for cross-development and testing). Therefore, we applied the general classification evaluation method to test the BICA Ding model. The development language was Python, and the packages called by the scheme included Sklearn,ango, Pandas, Pymysql, Spyder, etc. Therefore, online

Table 1. Data set information

Serial number	information	parameter
1	Number of resumes	61080
2	Number of positions	3000
3	Number of resumes accepted	33800
4	Number of resumes rejected	27280

recommendation methods in the recent 12 months are selected from some positions recommended by the company to collect resume samples accepted and rejected by the company. Positive samples are accepted and negative samples are rejected. Data was collected in Microsoft Office Excel format at the initial stage, and was imported into MySQL database format at the later stage. A total of 61080 samples were collected in the end, and the job categories included technology and design. The data set related information was shown in Table 1.

As can be seen from Table 1, the sample of resume accepted in the data set is that the enterprise has sent an offer to the job seeker, so we think the resume has been accepted. The rejected sample is the sample that the recruiter thinks the content of the resume does not match through the resume recommended by us. We collected 61080 resume samples, which were divided into training set and test set according to the ratio of 4:1. Under the above experimental preparation conditions, the experimental test is carried out and the experimental results are obtained.

3.2 Experimental Result

According to the needs of the experiment test, the selection is based on the Spark of human resource two-way recommended method, based on the human resources of the bidirectional circular convolution neural network recommended method, and the two-way recommended method through experiment contrast, human resources are tested under the condition of different number of users, three methods of recommendation accuracy, the larger the value, prove that the higher the accuracy. The calculation formula of the recommended accuracy rate is:

$$r = \frac{\sum_{g \in A} |H(g)| \cap |\eta(g)|}{\sum_{g \in A} |H(g)|} \times 100\% \tag{6}$$

In formula (6), g represents the user, H represents the number of tests in the test set, A represents the number of correct results in the first A items, and η represents the number of resumes. The experimental comparison methods are Spark-based bidirectional human resources recommendation method (method 1), and human resources bidirectional recommendation method based on recurrent convolutional neural network (method 2). According to the calculation results of Formula (6), the experimental results are shown in Fig. 2 and 3.
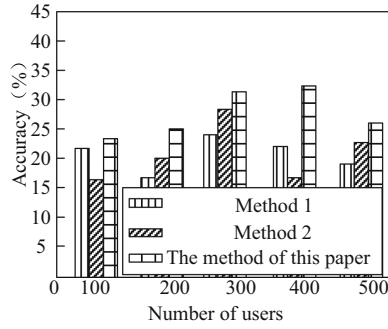
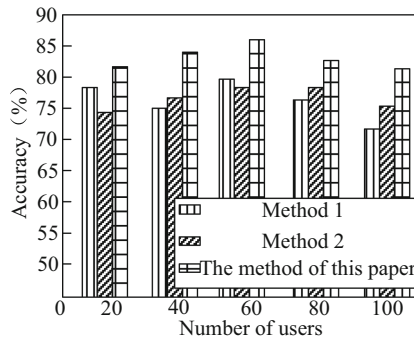**Fig. 2.** Number of users 500 recommendation accuracy (%)

Fig. 3. Number of users 200 recommendation accuracy (%)

According to Fig. 2, the average recommendation accuracy of the two-way human resource recommendation method in this paper and the other two two-way human resource recommendation methods is 27.842%, 20.312% and 20.618% respectively. According to Fig. 3, the average recommendation accuracy of the two-way human resource recommendation method in this paper and the other two two-way human resource recommendation methods are 83.448%, 76.414% and 75.915% respectively, and the two-way human resource recommendation method in the illustration has better performance.

4   Conclusion

In this paper, for the record of graduates sending resumes and the record of enterprises sending interview invitations, the matrix data is firstly constructed and then the matrix is improved by combining the two parts of information. Then, by constantly adjusting the proportion of each component to seek the optimal combination and adapt to different specific application requirements, the problem of data sparsity is solved by optimizing the data anonymous process to obtain the final matrix, and then the two-way recommendation for graduates and enterprises is realized. At present, the data of students and enterprises

obtained through data application and web crawler are not very comprehensive. In the future, I will continue to devote myself to relevant topics and make continuous progress.

References

1. Blazquez-Resino J J, S Gutiérrez-Broncano, Ruiz-Palomino P, et al. Dealing With Human Resources in the Age of Consumer 4.0: Aiming to Improve Service Delivery [J]. Frontiers in Psychology, 2020, 10:3058.
2. Gedam V V, Raut R D, Jabbour A, et al. Sustainable manufacturing and green human resources: Critical success factors in the automotive sector [J]. Business Strategy and the Environment, 2021, 30.
3. Kcc A, Chc A, Hht A, et al. Social media privacy management strategies: A SEM analysis of user privacy behaviors [J]. Computer Communications, 2021.
4. LU Li-hua, DU Cheng-lie. Multi-Layer Classification Storage Simulation of Complex Network User Privacy Data [J]. 2020, 37(3): 405-408,439.
5. Ronquillo J C, Popa A, Willems J. Toward an Understanding of the Role of Human Resources in Cultivating a Climate for Innovation in Nonprofit and Public Organizations [J]. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 2021:1-13.
6. Rusydi A R, Palutturi S, Noor N B, et al. The comparison of good corporate governance on human resources performance in regional general hospital of Labuang Baji and Ibnu Sina hospital Makassar, South Sulawesi [J]. Enfermería Clínica, 2020, 30:482-486.
7. Song M, Wang Z, Zhang Z, et al. Analyzing User-Level Privacy Attack Against Federated Learning [J]. IEEE Journal on Selected Areas in Communications, 2020, PP(99):1-1.
8. Stokes P. Human Resources [J]. Psychology Research, 2020, 10(10):3.
9. Singh N K, Mahajan V. End-User Privacy Protection Scheme from Cyber Intrusion in Smart Grid Advanced Metering Infrastructure [J]. International Journal of Critical Infrastructure Protection, 2021:100410.
10. Shashidhara R, Bojjagani S, Maurya A K, et al. A Robust user authentication protocol with privacy-preserving for roaming service in mobility environments [J]. Peer-to-Peer Networking and Applications, 2020, 13(3).
11. Thurman P, Michaels-Strasser S, Lukusa L, et al. Strengthening Human Resources for Health to Achieve Epidemic Control: A Case Study of Democratic Republic of the Congo [J]. SSRN Electronic Journal, 2020.
12. Vimalkumar M, Sharma S K, Singh J B, et al. 'Okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants[J]. Computers in Human Behavior, 2021, 120(3):106763.
13. Wu X, Wang H, Shi M, et al. DNA Motif finding Method without Protection can leak user Privacy [J]. IEEE Access, 2019, PP(99):1-1.
14. Wang Y, Zheng N, Xu M, et al. Hierarchical Identifier: Application to User Privacy Eavesdropping on Mobile Payment App [J]. Sensors, 2019, 19(14):3052.
15. Yamin M, Sen A. A New Method With Swapping of Peers and Fogs to Protect User Privacy in IoT Applications [J]. IEEE Access, 2020, 8:210206-210224.