ISSN 2616-7433 Vol. 2, Issue 7: 35-39, DOI: 10.25236/FSST.2020.020711

Research on User Identity Matching Algorithm of Online Social Network

Liu Qin

Shandong Institute of Commerce & Technology, Jinan Shandong 250103, China

ABSTRACT. *With the continuous progress of society and the rapid development of Internet technology, online social networks have become a hot topic and have attracted more and more registered users. In this situation, online social networks show a diversified development trend, diversified online social networks enrich people's online life, while social networks fragment personal information. The online social network has identity matching function, mainly for a person to exist multiple social network accounts at the same time, achieving account matching, and then complete the integration of information of online social network users, which not only plays a role in information association, product recommendations, but also plays an important role in network security. In response, this paper explores identity matching algorithms for structural information, spatio-temporal trajectory data and personal profile information in online social networks.*

KEYWORDS: *Online social networking, User identity matching, Algorithmic research*

1. Introduction

Because online social network user identity matching plays an important role in Internet Security and so on, there are many related researches, and great progress and development have been made. However, there are still shortcomings: (1) in combination with the current network structure information user identity matching algorithm research, for the higher dimension, higher-order structure information involved in the expression is not available, resulting in the final result The accuracy of the user identity matching algorithm is not high; (2) in the field of file information, there are subjective judgment and objective correction factors, and the weight allocation method has subjective factors that do not take into account the implicit relationship between attributes, resulting in low accuracy; (3) the main feature of the user's spatiotemporal trajectory information is to accurately describe the spatiotemporal trajectory, which is in the current user identity matching algorithm. It is difficult to describe the spatiotemporal trajectory because of the lack of the ability to extract the features of location access sequence. In view of the above shortcomings, this paper proposes corresponding user identity matching algorithms for the network structure information, user spatiotemporal trajectory data and user profile information in online social networks.

2. User Identity Matching Algorithm Based on Network Representation Learning

2.1 Basic Content

Network structure information: network structure information is the friend relationship between users of social networks. When users first contact with social networks, they often establish a social circle in real life, which maps real life to the network environment. To some extent, network structure reflects the social relationship of network users, so the network structure information of social networks is open transparent, and there is no false information, data. For this real, open data information, network structure information has other insurmountable advantages. Based on it, more researchers focus on how to mine network structure information, and through nodes to match the user identity of network structure characteristics has become a hot topic.

In the past research and analysis of user identity matching of network structure information, it is found that many algorithms are based on the similarity of nodes, and calculate the similarity of cross network nodes according to the similarity index of nodes. This algorithm analysis is simple, without in-depth analysis of network structure, and in the case of dimension limitation and complexity of calculation, only the adjacent nodes can be determined. For example, the first or second order, deeper and more complex cross network nodes cannot be accurately calculated. In addition, there will be nodes with die of network structure features in large-scale data. With the increase of nodes, the accuracy of this algorithm decreases, so it is difficult to accurately calculate and

ISSN 2616-7433 Vol. 2, Issue 7: 35-39, DOI: 10.25236/FSST.2020.020711

determine matching accounts. More importantly, this algorithm lacks the assistance of other dimension information, so it will affect the calculation performance when facing similar nodes in the network structure. There are also many researchers with the help of user information do not involve privacy attributes. The combination of user attributes and network structure information, the use of user information to obtain and data reliability strong characteristics, the integration of multi-dimensional information to match user identity, although this method has some effect, the lack of in-depth analysis of network structure, calculation method still has room to improve.

Combined with the shortcomings of the previous methods of identity matching calculation, this paper proposes a user identity matching algorithm for network representation learning based on network structure information. The main direction is to combine the user's attribute information for user identity matching calculation. Based on the network structure information, the network representation learning technology is simpler than the traditional calculation method, and it is difficult for the traditional calculation method. In the calculation of the high-order node information, the network representation learning technology can be competent to complete. On the basis of this improvement, the user name attribute information is integrated, the network structure information and the user name information account are combined, and the matching account is judged by the vector similarity. According to the real practice results, the comprehensive performance of the network representation learning technology algorithm is high.

2.2 Algorithm Flow

Through the fusion vector representation of network structure information and user name information nodes, the similarity between the nodes across the network needs to be calculated. Calculation method: firstly, the nodes in the network are distinguished according to their attribute networks, which are ς^X nodes and ς^Y nodes, and then find cosine similarity in the fusion information vector of cross network nodes, and take cosine similarity as the similarity between accounts. When the ς^X node to be matched u_i^X appears in the network, when the u_j^Y node is the account ς^Y node with the highest similarity u_i^X node in the network, it is called the u_j^Y candidate matching account u_i^X. This algorithm adopts two-way matching strategy, that is, only u_i^X and u_j^Y two directions, which are the candidate matching accounts of each other. But only when the similarity is greater than the matching threshold T, can we conclude that the two accounts are matched and form the seed accounts together. If the similarity does not meet the matching threshold T, we need to re match the nodes.

3. User Identity Matching Algorithm of File Information Fusion Based on Fuzzy Integral Theory

3.1 Basic Content

User profile information: the user profile information mainly includes name, gender, age, hobbies, address, nickname, graduation school and work resume, etc., so the user profile information can comprehensively reflect the basic personal information of social network users. In order to make use of various attribute information in the user profile and match the similarity of social network accounts, it is necessary to ensure the profile information. In order to ensure the accuracy of network user identity matching.

Aiming at the user identity matching of multi-attribute user's file information, many researchers have made a breakthrough in the algorithm of multi-attribute file information through long-term exploration and practice. Through the research on the calculation method, we find that most of the calculation ideas are basically to calculate the similarity of individual attributes, determine the user identity according to the attributes, and assign the user attributes to them. Finally, the weights are compared and fused to get the final account similarity result and complete the final matching decision.

In the above calculation methods, the similarity of attributes is mainly measured by character similarity, and the weights of attributes are usually determined by subjective experience. The modification of weights is also calculated by matching results. Although this method has been applied in practice, there are still some problems: (1) the similarity of attributes is judged by character similarity, although The judgment is simple, the algorithm is not complex, but there are differences between individual attributes, and some attributes have special properties, which is not taken into account in this method, resulting in differences in similarity; (2) the subjective weight value of this method is generalized, and the attributes may be different in different social networks, and the same attribute is in different social networks The importance degree in may also be different, so the general subjective

ISSN 2616-7433 Vol. 2, Issue 7: 35-39, DOI: 10.25236/FSST.2020.020711

weight value cannot meet the occurrence of more situations.

In order to improve the disadvantages of the above calculation methods, this paper puts forward the concept of fuzzy integral for the user identity matching algorithm of file information fusion. The basic problem is the file information, and the purpose is to solve the accurate user identity matching. The fuzzy integral theory can reasonably allocate the weight, realize the similarity judgment of different situations and different attributes in the traditional way, and the calculation method will discrete the integral. Combined with the theory of fuzzy measure, according to the different attributes in different file information, we allocate them, and then calculate the similarity. The basis of attribute weight comes from the particle swarm optimization algorithm. Finally, we use the integration to compare the similarity between accounts. According to the practice results, it shows that the efficiency is improved compared with the traditional calculation method.

3.2 Algorithm Flow

The integration of fuzzy measure theory and multi-attribute information on-line user identity matching mainly includes the following steps: (1) attribute similarity calculation: the account to be matched is calculated to get attribute similarity; (2) fuzzy density calculation: the account to be matched in the case of fuzzy density is similarly scored; (3) fuzzy density optimization: the fuzzy density optimization is calculated by PSO, The corresponding attribute fuzzy density of each account to be matched, the combination and fuzzy measurement of each account to be matched; (4) integral calculation and account matching: finally, set the matching threshold T according to the above score and calculation for comparison, so as to determine whether the account is matched or not, as shown in the summary, the detailed matching algorithm is as follows:

Table 1 1 User Identity Matching Algorithm Based on Fuzzy Integral Theory

Algorithm 5: user identity matching algorithm based on fuzzy integral theory

Input: social networks $\mathcal{G}^X=(\mathcal{U}^X,\mathcal{P}^X)$ and $\mathcal{G}^Y=(\mathcal{U}^Y,\mathcal{P}^Y)$,
seed accounts set $\mathcal{S}$,
similarity threshold T

Output: final matching accounts set $\mathcal{S}'$

```
for each (u_i^X, u_j^Y) from G^X and G^Y :
    calculate Sim(p_i^X, p_j^Y) according to Sec 4.3
end for
calculate each g_i according to algorithm 4
calculate λ by eq.(4.4)
calculate each g_λ(U_i) by eq.(4.5)
for each (u_i^X, u_j^Y) from G^X and G^Y :
    calculate C_μ by eq.(4.6)
    if score > T:
        add (u_i^X, u_j^Y) to S'
    end if
end for
return S'
```

4. User Identity Matching Algorithm Based on Sequential Feature Representation of Spatiotemporal Trajectory

4.1 Basic Content

With the continuous progress of science and technology and the rapid development of Internet industry, mobile Internet devices have become necessary and popular. The functions of social networks on the mobile end are more and more diversified. The function of sharing one's own location is also added to the published content. Some social networks are functional needs, and they will automatically save the user's real-time location, thus forming a space-time orbit A lot of data of trace, so the spatiotemporal trace also reflects the user's action trace in real life, so more researchers can mine the information in the spatiotemporal trace data, which can realize modeling, location determination, location prediction, etc., so the spatiotemporal trace is a reflection of human's

ISSN 2616-7433 Vol. 2, Issue 7: 35-39, DOI: 10.25236/FSST.2020.020711

relevant characteristics, and can use the spatiotemporal trace information to complete the identity matching of online social network users.

According to the previous use of spatiotemporal trajectory data to complete user identity matching, it is found that most of the calculation methods are based on frequency, through the frequency of fixed location trajectory access, or the frequency of two overlapped tracks to calculate similarity. Although this method has achieved results in practice, there are still the following problems: the mining of access sequence features is not sufficient, but simple Considering spatiotemporal trajectory data as coordinate points, the result of identity matching is not convincing, and this way is limited by dimensions, and the direction of consideration is not comprehensive and sufficient.

In view of the above problems, this paper explores the sequence characteristics of spatiotemporal trajectory, and completes the calculation method of user identity matching. The calculation method mainly uses the structure of cyclic neural network to achieve the problem of poor extraction ability in the traditional spatiotemporal trajectory matching settlement method. Firstly, the track data is transformed into grid form for preprocessing, and the track granularity is segmented according to the time granularity. Then, the track points are mined by using the relevant methods. Finally, the track sequence is extracted by using Bi-GRU model, from which the features of position access sequence are mined, and the track vector is calculated, and the account matching is completed according to the track vector.

4.2 Algorithm Flow

(1) User track matching

Firstly, the trajectory is set to T, then the nodes in T are vectorized and input to the preset network longitudinal in order, and finally the trajectory vector T_i is obtained. For different trajectories in two networks, such as T_i^A T_i^B vector sum, we can use cosine similarity of vector to calculate, and compare with threshold value, according to this result, we can judge whether two trajectories come from the same user in reality.

4.3 Algorithm Flow

Table 2 1 User Identity Matching Algorithm Based on Spatiotemporal Trajectory Sequence Feature

ISSN 2616-7433 Vol. 2, Issue 7: 35-39, DOI: 10.25236/FSST.2020.020711

Algorithm 6: user identity matching algorithm based on trajectory representation

Input: social networks $\mathcal{G}^X=(\mathcal{U}^X,\mathcal{T}^X)$ and $\mathcal{G}^Y=(\mathcal{U}^Y,\mathcal{T}^Y)$,
seed accounts set $\mathcal{S}$,
interval of latitude and longitude γ and φ,
dimension of the location vector d,
learning rate α,
similarity threshold S

Output: final matching accounts set $\mathcal{S}'$

```
data preprocessing according to Sec.5.2
Initialize location vectors
for each c_i in C:
    for each trajectory T including location c_i:
        calculate T according to the Bi-GRU model
        update parameters by eq.(5.13)
        update c_i by eq.(5.6)
    end for
end for
for each pair of trajectories maintaining the matching condition in Sec.5.5.1:
    calculate the similarities between the two trajectories s(T_i^X, T_j^Y)
    if s(T_i^X, T_j^Y) > S:
        add (u_i^X, u_j^Y) to S'
    end if
end for
return S'
```

Representation

The main process of Bi-GRU model in user trajectory matching method is data preprocessing, vectorization and other steps. According to the trajectory vector, it completes the identity matching of online social network users.

5. Summary

With the progress of science and technology and the advent of the Internet era, more and more users use social networks to communicate, entertain, socialize, work and so on. As a result, more and more users have multiple social network accounts at the same time. Therefore, online social network user identity is realized Matching has become an important content. In terms of scientific research, business and network security, user identity matching is of great significance and value. This paper studies the online social network Yonghua identity matching calculation method, hoping to provide advice and help to the Internet industry.

References

[1] Zhang Shusen, Liang Xun, Mi Baotong, et al (2019). Content based social network user identification method. Journal of computer science, vol.42, no.8, pp.1739-1754.
[2] Yang Yizhuo, Yu Hongtao, Huang Ruiyang, et al (2018). Cross social network user identity matching based on fusion representation learning. Computer Engineering, vol.44, no.9, pp. 45-51.
[3] Wu Zheng, Yu Hongtao, Huang Ruiyang, et al (2018). Cross network user identification based on hidden label node mining. Computer application research, vol.35, no.4, pp.1191-1196.
[4] Wu Zheng, Yu Hongtao, Liu Shuxin, (2017). Method of cross social network user identification based on information entropy. Computer application, vol.37, no.8, pp.2374-2380.