

Research Article

A Stable-Matching-Based User Linking Method with User Preference Order

Xuzhong Wang, Yan Liu, and Yu Nan

China State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450002, China

Correspondence should be addressed to Yan Liu; ms_liuyan@aliyun.com

Received 10 December 2016; Revised 18 February 2017; Accepted 1 March 2017; Published 28 March 2017

Academic Editor: Zonghua Zhang

Copyright © 2017 Xuzhong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of social networks, more and more users choose to use multiple accounts from different networks to meet their needs. Linking a particular user's multiple accounts not only can improve user's experience of the net-services such as recommender system, but also plays a significant role in network security. However, multiple accounts of the same user are often not directly linked to each other, and further, the privacy policy provided by the service provider makes it harder to find accounts for a particular user. In this paper, we propose a stable-matching-based method with user preference order for the problem of low accuracy of user linking in cross-media sparse data. Different from the traditional way which just calculates the similarity of accounts, we take full account of the mutual influence among multiple accounts by regarding different networks as bilateral (multilateral) market and user linking as a stable matching problem in such a market. Based on the combination of Game-Theoretic Machine Learning and Pairwise, a novel user linking method has been proposed. The experiment shows that our method has a 21.6% improvement in accuracy compared with the traditional linking method and a further increase of about 7.8% after adding the prior knowledge.

1. Introduction

Over the past decade, followed by the exponentially growing net-services, the number of anonymous users is also springing up. As of the third quarter of 2016, active users of Facebook reached 1.79 billion [1], which means more than half of 30 million Internet users use Facebook per month at least once. About 65% or about 1.18 billion users log at least once in daily. However, some traditional social network sites now are facing significant development. According to the Twitter 2016 Q3 results [2], average growth rate of monthly active users, only about 3%, reached 317 million, compared with the image-based social network Instagram, whose monthly active users have already exceeded 600 million [3]. This change shows that, with the development of times, user's interest of the net-services has been divided. Therefore, net-services providers also aim at developing different social services for various user's interests.

Nowadays, each net-service often has its unique mode of information sharing to maintain its social relationships. These unique models attract different user groups; for

example, a user selects Twitter to share some information publicly and chooses Facebook for own circles, and for sharing traveling scenery and foods, of course Instagram is the best choice. On these net-services platforms, users typically pass a uniquely identified nickname along with some other attribute tags, such as profile information, hobbies, friendships, and events. If these accounts can be effectively linked with a particular user, when we try to understand a user comprehensively, this not only can significantly improve his (or her) experience of a recommender system but also can provide a better anonymity protection policy [4]. In network security, when detecting malicious attackers with multiple accounts in different platforms, it is possible to integrate the information of cross-media together and makes a vast improvement of the detecting ability. Practice has proved user linking has important practical significance.

However, due to the anonymity protection policy of net-services providers and users in different net-service platforms always choose to share different information, resulting in the fact that particular user's multiple accounts often do not have

adamant relevance. This large number of nondirectly linked account makes difficulties for comprehensively understanding the user. There are existing studies done by analyzing the user's naming style convention [5, 6], profile [7, 8], writing style [9], behavior [10], social relations [11, 12], and so on and then by linking users multiple accounts by statistical and also machine learning methods. These methods are used to model the characteristics of a vast number of accounts and made some certain achievements in the experimental dataset. However, in reality, not enough account features could be gotten from sparse network data and the behavior behind these accounts is always changing. It is hard to use a stable mathematical model to describe it. Moreover, the real human behavior is neither random nor entirely rational [13]. Therefore, considering the mutual influence among multiple accounts from different net-services users, the user linking problem can be regarded as a cooperative game problem in the bilateral (multilateral) market—how to formulate a cooperation (linking) strategy in the markets (net-services) to enhance the interest (linking results) of the whole candidate accounts set.

In recent years, researches on Game-Theoretic Machine Learning are progressing; some researchers have constructed a Game-Theoretic Machine Learning framework, through the Markov model to study and predict the user's behavior [13–15]; some scientists use cooperative game approach to evaluate and select the features for machine learning [16]. These methods have proved the game theory plays an improving role on the traditional machine learning. Therefore, this paper proposes a stable-matching-based game theory method for user linking with user preference order and prior knowledge. The main contributions of this paper are as follows.

Process a novel method based on stable matching game theory to carry on the analysis of user linking.

Input the linked user accounts as prior knowledge to enhance the result of user linking.

Through many experiments carried out in the LifeSpec [17] project dataset provided by Microsoft Research Asia, our method is about 21.6% higher in accuracy compared with the traditional user linking methods. Moreover, there was a further improvement of about 7.8% after inputting the prior knowledge.

2. Problem Formulation

In this section, the related concepts and formal descriptions of user linking are given. For the convenience of description, this paper focuses on two heterogeneous networks. Symbols used in this paper are shown in Table 1.

User. A user can be represented as $u_i = \{a_i^1, a_i^2, a_i^3, \dots, a_i^n\}$, where $a_i^k = \{f_1, f_2, \dots, f_p\}$ represents the user u_i 's account on the network k and f represents feature of accounts. For convenience, we focus on two heterogeneous networks, so $u_i = \{a_i^s, a_i^t\}$, where a_i^s represents the account from the source network s and a_i^t represents the account from the target network t .

TABLE 1: Symbols used in this paper.

Symbols	Significance
u	User
a	User account
A	Account set
pair	Accounts pair
l_{pair}	Identification of accounts pair
s	Source network
t	Target network

Account Set. An account set represents extractable accounts from a particular network. So $A^s = \{a_1^s, a_2^s, a_3^s, \dots, a_p^s\}$ represents source network account set and $A^t = \{a_1^t, a_2^t, a_3^t, \dots, a_q^t\}$ represents target network account set, where p, q are the number of users of both networks.

Accounts Pair. Accounts pair pair = (a_m^s, a_n^t) represents a tuple consisting of any account a_m^s of user u_m from the source network s and any account a_n^t of user u_n from the target network t .

Identification of Accounts Pair. $l_{\text{pair}=(a_m^s, a_n^t)} = \{0, u_m \neq u_n; 1, u_m = u_n\}$, which means when the accounts pair consisted of accounts from the same user, the value of identification $l_{\text{pair}} = 1$; otherwise $l_{\text{pair}} = 0$.

Problem Description. Given the source network s and target network t , extracting the candidate account sets A^s, A^t , and grouping any two accounts from these networks one by one to an accounts pair, then get $n * m$ pairs. Finally, use a linking algorithm to find all the pairs whose identification $l_{\text{pair}} = 1$, namely, linking accounts a_i^s, a_j^t from two heterogeneous networks s, t .

The challenges of this paper are as follows:

- (1) Traditional user linking technology is often trying to maximize some objective function so that the whole candidate accounts set can get the best result. However, since the user's different account behavior is often not rational and stable [13] and the sparse features of accounts could influence the linking result significantly, the traditional methods do not always have an ideal result on large-scale sparse data sets. Within the cooperative game theory, user linking is actually trying to find matched players in the bilateral market. In this paper, we combined the game theory and the user's preference using stable matching theory [18] and Pairwise, finally linking users through the cooperation between accounts.
- (2) The traditional method often linked the user's account by calculating the "similarity" between different accounts using certain types characteristics. However, in the real world, multiple accounts of a user on various platforms tend to reflect different needs of the user, resulting in the fact that the "similarity" is

minuscule that many accounts can not be linked. Taking into account the fact that the “user linking problem” and “linking similar user problem” are different, so we input some linked user as prior knowledge, thereby enhancing the result of user linking.

The following section will detail how to solve these two problems.

3. Stable User Linking with User Preference Order

User linking essentially is a multiclassification problem, and different user accounts are categorized according to the user category. However, because the multiclassification problem is usually difficult to obtain an ideal solution, therefore, in this paper, we make use of the idea of Pairwise [19], combined accounts to pairs, and classified them according to whether linked. Then user linking problem will be converted into a binary classification problem and could calculate the probability of each account pair under a different category. Then, according to this probability, construct the user preference order set and finally convert the question into “how to select the best target account in one’s preference order set” and try to improve it by inputting the prior knowledge. Therefore, we present a three-phase approach to solve the user linking problem:

- (1) Constructing user preference order set: calculating posterior probability P for each pair according to the SVM model trained by the training set and sorting P of each a^s to construct user preference order set.
- (2) User linking based on stable matching: using stable matching algorithm based on the user preference order set between A^s, A^t and finally getting all the stable links among accounts.
- (3) User linking based on prior knowledge: inputting the prior knowledge to improve user linking in the stable matching algorithm and finally get the reinforced user linking algorithm.

3.1. Constructing User Preference Order Set. According to Pairwise, first user linking can be converted into a binary classification problem, and by calculating the classification probability the account preference order set could be constructed, which is defined as follows.

User Preference Order Set. For an account a^s , the ordered sequence $\{a_1^t > a_2^t > a_3^t > \dots > a_n^t\}$ of the target account set A^t is called the preference order set of the account a^s . The ordered sequence reflects the order of which target account a^s is more likely to link.

In recent years, many kinds of research have shown that the Support Vector Machine has a high ability in resolving the problem of binary classification [20, 21]. Since SVM is very sensitive to features, selecting the proper feature is vital. Traditional methods make many features by artificial information, such as naming habits, personal profiles, writing

style, user behavior trajectory, and social relations. However, due to the incompleteness and heterogeneity of network data, the features of user data acquired not only can be very limited but also need to be completed. Therefore, by using account labels, we avoid the difficulty of filtering and completing of features.

From the reality net-services, some of them provide labels to simply and clearly reflect the characteristics of user accounts. But others do not have. So, we can directly construct labels by account history text using topic model, such as LDA. The method of label extraction by the topic model has been matured in recent years and will not be repeated here.

In this paper we took these accounts with their label tag as a bag-of-words model and then calculate the value of features between a_m^s 's feature vector $T_m^s = \{\text{tag}_1, \text{tag}_2, \dots, \text{tag}_p\}$ and a_n^t 's feature vector $T_n^t = \{\text{tag}'_1, \text{tag}'_2, \dots, \text{tag}'_q\}$ as follows:

$$(1) \text{ Cosine similarity: } \cos = (T_m^s \cdot T_n^t) / \|T_m^s \times T_n^t\|.$$

$$(2) \text{ Number of common labels: } n = \#(T_m^s \cap T_n^t).$$

According to the feature above, the training data can be trained by SVM and then accurately classify the test data. However, in large-scale data, there are many accounts because of the sparseness of labels and the different user’s accounts may have some similarity, resulting in the fact that many cases can not make an accurate classification. These noise accounts will have a great impact on the classifying effect when using the standard SVM. In fact, user linking is a nondeterministic classification problem: some samples can not belong to a category accurately, only through the probability to reflect its belonging to a certain category. To address this issue, according to the sigmoid-fitting method proposed by Platt [22], we calculate each pair’s posterior probability P under the conditions $l_{\text{pair}} = 1$:

$$P(l_{\text{pair}} = 1 | f) = \frac{1}{1 + e^{Af+B}}, \quad (1)$$

where f is the Support Vector Machine with no threshold output $f(x) = w^t x + b$ and two parameters A, B can be set by maximum likelihood estimation of the training set. This posterior probability actually reflects the likelihood that one account will be linked to another target account. According to the posterior probability we construct user preference order set as follows.

Based on Pairwise, the training set and the test set of pairs are constructed between account sets A^s, A^t , and the feature vectors of any pairs are constructed by using the above two features, and then use Support Vector Machine to train a model on the training set. For a particular test set account a^s , calculate the posterior probability P of pair = (a^s, a^t) , where a^t comes from target network t , under the conditions $l_{\text{pair}} = 1$. Finally, we get the user preference order set $\{a_1^t > a_2^t > a_3^t > \dots > a_n^t\}$ of a^s by sorting P of each pair.

The following section describes how to link user accounts by user preference order set.

```

Input: account set  $A^s, A^t$ 
Output: result set  $T$ 
(1) Initializes the result set  $T = \emptyset$ 
(2) Calculate the posterior probability  $P$  of any pair  $= (a_s, a_t)$ 
(3) Sort to get the preference order set for each account
(4) if  $A^s \cdot \text{length}() \neq A^t \cdot \text{length}()$  then
(5)   Add  $A^s \cdot \text{length}() - A^t \cdot \text{length}()$  fake accounts to the small parties, and set the preference order set keeps empty
(6) end if
(7) while Exists any account  $a_s \in A^s$  is not linked &&  $a_s$ 's preference order set  $\neq \emptyset$  do
(8)   Find the most preferred target account  $a_t$  from  $a_s$ 's preference order set and remove it
(9)   if  $a_t$  is not linked &&  $a_s$  is in the preference order of  $a_t$  then
(10)     Set  $a_s, a_t$  linked,  $T = T \cup \{(a_s, a_t)\}$ 
(11)   else if  $a_t$  is linked then
(12)     Get the linking object  $a_m$  of  $a_t$ 
(13)     if  $a_s$  is in the preference order set of  $a_t$  &&  $a_s > a_m$  then
(14)       Cancel the linking state of  $a_m$ ,  $T = T - \{(a_m, a_t)\}$ 
(15)       Set  $a_s, a_t$  linked,  $T = T \cup \{(a_s, a_t)\}$ 
(16)     end if
(17)   end if
(18) end while
(19) Remove all the accounts linked with  $a_f$ 
(20) return  $T$ 

```

ALGORITHM 1: Stable User Linking with Preference order, SULP.

3.2. User Linking Based on Stable Matching. Through the convention in Section 3.1, user linking actually turns into “how to select the best target account in one’s preference order set” so that the whole candidate account set can get the best performance. In this paper, we try to use stable matching theory to solve this problem. The stable matching theory [18] is proposed by Shapley using cooperative game theory to solve the linking problem in bilateral market entities. Because of this theory, Shapley won the 2012 Nobel Prize in Economics. This theory has been widely used in many practical scenarios, such as students selecting (students and schools matching [23]), housing allocation (matching between people and house [24]), and job searching (employee and employer matching [25]). The core of this theory lies in the realization of the stable state, which means there does NOT exist ANY pair of entities in the bilateral market at the end of linking, which have a more preferred target than the currently linking target. In fact, if the source network s and target network t are regarded as a bilateral market, user accounts a_m^s, a_n^t can be seen as entities from the bilateral market. Then the problem of “how to select the best target account in one’s preference order set” is converted to “how to find a cooperation (linking) strategy in the markets (networks) to make the interest (linking results) to the maximum.” Therefore, based on the idea of stable matching, we linked accounts based on the preference order set.

Broken Account Pair. If an account a_m^s is linking to a_n^t , a_n^t is linking to $a_{n'}^s$. Assume there is a pair $= (a^s, a_n^t)$ on which the account a_m^s has $a_n^t > a_{n'}^s$ in its preference order set and the account a_n^t has $a_m^s > a_{n'}^s$ in its preference order set; then the pair $= (a^s, a_n^t)$ is called a broken account pair because actually it breaks the current linked pairs.

Stable Matching. If there does NOT exist ANY broken account pair at the end of linking, then we said the entire linking is a stable matching.

Using [18] proposed GS delay algorithm can achieve a stable matching in the bilateral market. However, the standard GS algorithm requires that the number of entities in the bilateral market must be N , and the size of preference order set of each entity must also be the size N . That is to say, “the number of bilateral market entities is same” and “each preference order set is completed.” However, these two restrictions are difficult to meet, and because of the lack of attributes, some of the feature vectors can not be calculated and can not get the completed order set, so we make two adaptations.

- (1) Fake account: an account which does NOT actually exist is called fake account a_f . In a linking process, a balanced number between two account sets of fake accounts will be added to the littler set, and when linking is completed all the pairs which contain fake account will be excluded.
- (2) Uncompleted user preference order set: a user preference order set which does NOT include ALL the accounts in the target network is called an uncompleted user preference order set. In a linking process, if a^t is not in a^s 's user preference order set we directly denied this link.

According to this, we propose a stable-matching-based user linking method with user preference order (Stable User Linking with Preference order, SULP) as shown in Algorithm 1.

Through Algorithm 1, this paper combines the user preference order and stable matching of cooperative game

```

(1) the same as SULP Algorithm 1 line 1–8
(2) if  $a_t$  is not linked &&  $a_s$  is in the preference order of  $a_t$  then
(3)   Set  $a_s, a_t$  linked,  $T = T \cup \{(a_s, a_t)\}$ 
(4) else if  $a_t$  is linked then
(5)   Get the linking object  $a_m$  of  $a_t$ 
(6)   if  $a_s$  is a prior candidate account of  $a_t$  then
(7)     if  $a_m$  is not a prior candidate account of  $a_t \parallel a_s < a_m$  then
(8)       Cancel the linking state of  $a_m, T = T - \{(a_m, a_t)\}$ 
(9)       Set  $a_s, a_t$  linked,  $T = T \cup \{(a_s, a_t)\}$ 
(10)      Clear all the accounts behind  $a_s$ 's preference order list
(11)     end if
(12)   end if
(13) end if
(14) Remove all the accounts linked with  $a_f$ 
(15) return  $T$ 

```

ALGORITHM 2: EXTENDED Stable User Linking with Preference order, EXSULP.

theory to achieve the purpose of user linking. The next section will be on how to strengthen the result of this method.

3.3. User Linking Based on Prior Knowledge. Consistent with the traditional linking method, the method we proposed is still based on the similarity of account features. However, in fact, as the network platform tends to specify functionally, users on different platforms usually choose to explicitly express their interest by their multiple accounts, and these various interests among the accounts are likely to have little similarity. Therefore, user linking not only is “how to link accounts by similarity,” but also includes “how to identify and link the accounts which are dissimilar but belong to the same user.” The latter one is extremely challenging, and the researches show that there has been no effective solution. In this paper, we try to input some users’ linked accounts as prior knowledge, to strengthen the user linking method proposed in Section 3.2.

Considering that the preference order set of the entity in the bilateral market is a set based on the feature similarity, the above method can not adequately reflect the correlation information among different accounts. To add some correlative information by prior knowledge, we defined prior candidate account set as follows.

Prior Candidate Account Set. For an account a^s , given its linked account a^t , then a^t is called a prior candidate account of a^s . In the matching process, a^s is assumed to match account $a^{t'}$, if $a^{t'}$ is NOT a prior candidate account of a^s ; then regardless of the preference order set, let $a^{t'}$ link to a^s . If $a^{t'}$ IS a prior candidate account of a^s , then follow the order of preference set.

Based on the definition above, we further propose a reinforced algorithm (EXTENDED Stable User Linking with Preference order, EXSULP) based on prior knowledge. Only the improved part is shown in Algorithm 2.

According to the algorithm, we input the already linked account as the prior knowledge, further strengthening the

TABLE 2: My second table.

Dataset	Number of accounts	Number of comments	Number of works
Books	34942	2118400	523064
Movies	41823	8397846	82868

possible correlation between the accounts. Finally, all the eligible pairs = (a_m^s, a_n^t) are taken as the final result of user linking between the network s and network t .

4. Experiments

In this section, based on the dataset provided by Microsoft Research Asia LifeSpec [17], we used the standard SVM, SVM based on the cooperative game theory, and reinforced SVM based on prior knowledge, respectively, to analyze user linking. Experiment code has been made public on GitHub: <https://github.com/Observerspy/UserStableMatching>.

4.1. Dataset Description. LifeSpec is a computational framework developed by the Microsoft Research Asia for discovering and hierarchically categorizing urban lifestyles. The LifeSpec dataset is composed of tens of millions of user’s data about sign-in, movie comments, book comments, music comments, and behavior. In this paper, we attempt to link users from the books set as the source network s and to movies set as the target network t .

As in Table 2, we selected a total of 62,558 different users.

- (1) Books Dataset: contains 34,942 different accounts on 523,064 books with 2,118,400 comments; each data contains title, author, publisher, date of issue, number of pages, price, packaging, labels, user ratings, and other information.
- (2) Movies Dataset: contains 41,823 different accounts on 82,868 movies with 8,397,846 comments; each data contains name, director, screenwriter, starring,

category, country, duration, release date, labels, user ratings, and other information.

The total number of pairs in this dataset is 1,461,379,266. Because, in such a large-scale dataset, the proportion of positive instances and negative instances is often more than 1:10000, we controlled the proportion to about 1:1 by random undersampling.

4.2. Performance of User Linking Methods. We took labels from the books and movies as the accounts features and the frequency of each label as the feature value. Because the dimension of inputting feature vector is small, we use ten times 10-fold cross-validation Gaussian kernel SVM with setting the cost value to 1 and remaining the default parameters. Support Vector Machines and posterior probability calculations are provided by LibSVM [26] tools. The compared methods are summarized as follows.

- (1) SVM_Label: baseline method, using SVM to do a link\nonlink classification only in label feature space.
- (2) Sulp: the stable-matching-based user linking method with user preference order which is proposed in Section 3.2.
- (3) EXSulp: the extended user linking method which is proposed in Section 3.3.

As the user linking problem only concerned with the correct links (positive instances), therefore, we select the

precision p , recall r , and $F1$ value $F1$ as the evaluation metrics, and the average result of 10 times 10-fold cross-validation is shown in Table 3.

It can be seen from the results that the two methods proposed in this paper have surpassed the baseline method on the metrics of precision p , recall r , and $F1$, where the Sulp has an improvement of about 21.6% in accuracy and a further increase of about 7.8% after adding the prior knowledge. Compared with other researches which used a large number of user's personal information, texts, behaviors, and so on, we achieved the ideal precision when only using the labels as a feature. Moreover, different from other stable matching methods [27], we canceled the two restriction conditions of the following: "the number of bilateral market entities must be same" and "the preference order set is completed." Therefore, in the complex sparse real dataset, the method proposed in this paper can be considered to have better practical significance.

4.3. Analysis of Prior Knowledge. From the experiment above, we can know that the prior knowledge can improve the performance of user linking. It is clear that the proportion of prior knowledge to the whole data will influence the final linking results. Therefore, we analyze EXSulp algorithm by taking a part of incorrect classification results (total 2158) obtained from Sulp algorithm as a prior knowledge and changing the proportion of the prior knowledge to analyze the effect of prior knowledge.

Expansion Rate

$$\text{Expansion rate} = \frac{\#(\text{EXSulp right classification}) - \#(\text{Sulp right classification})}{\text{a priori knowledge's proportion of total incorrect instances}}, \quad (2)$$

representing the extended ability of the EXSulp algorithm for linking results.

The result is shown in Figure 1.

From the results, it can be seen with the increasing proportion that p , r , and $F1$ values increase steadily. It can be considered the proportion of prior knowledge is in proportion to the result of the algorithm, enhancing the precision of up to about 7.8%. The expansion rate reflects the fact that the results of this algorithm gradually stabilize as the scale of prior knowledge increases. The above experiment sufficiently proved the prior knowledge can enhance the correlation among accounts, illustrating the effectiveness of our method.

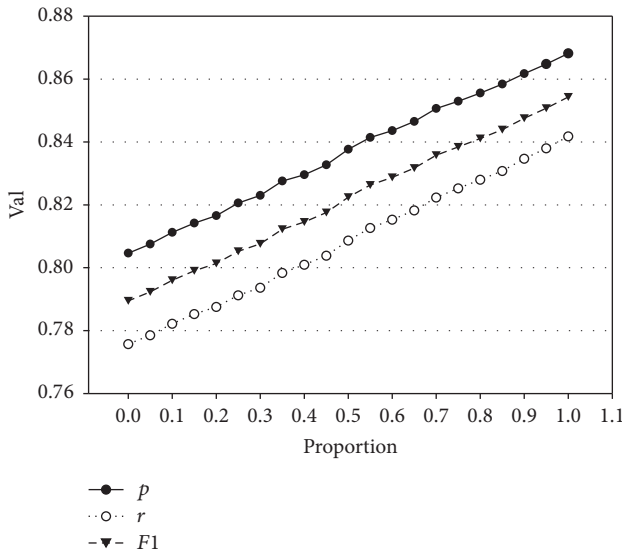
4.4. Case Study. We choose four linked results to display and analyze in Table 4. The coexisting top 10 labels are given (translated to English, the works name is in italic), among which 1–3 are the correct links and 4 is a wrong link.

As can be seen from Table 4, because of the semantics of the label, when the coexisting labels are specific enough, then the accounts can be correctly linked. In fact, it further illustrates that the problem of user linking can be solved

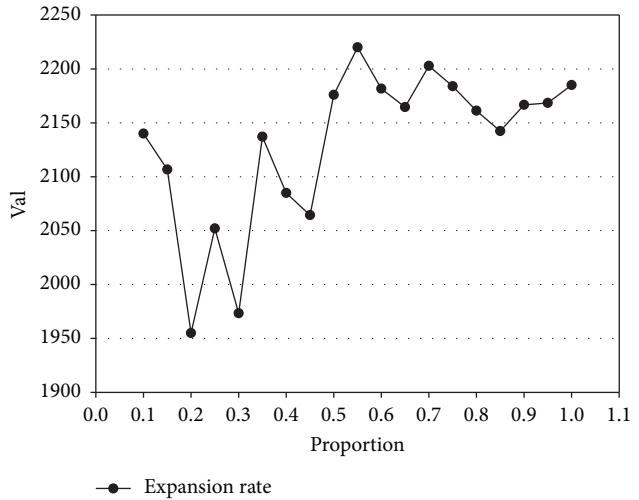
according to each user's specific and unique interest labels. However, as shown in item 4, when these labels represent more abstract and general terms, these accounts can not make the right link. When the label of the linked account which is inputted as the prior knowledge contains such abstract and general terms, it can effectively reduce the misclassification caused by the classifier based on calculating features.

5. Conclusion

In this paper, we have studied the user linking problem and propose a stable-matching-based method with user preference order. Different from the restrictions of the traditional stable matching algorithm, we made some relaxation and enhance the result of user linking by inputting prior knowledge. Experiments show that, in the real dataset, our method has achieved an ideal effect when only using the characteristics of the website label, which adequately demonstrates the effectiveness of this approach. In the future research, we will further study how to extract accurate and efficient characteristics in the sparse data and how to enhance the correlation between different accounts.



(a) Effect on the p , r , $F1$ values



(b) Effect on the expansion rate

FIGURE 1: Effect of CGSVMEX on the proportion of prior knowledge.

TABLE 3: Performance of different methods.

Methods	p	r	$F1$
EXSULP	86.8%	84.2%	85.5%
SULP	80.5%	77.6%	79.0%
SVM_Label	66.2%	40.9%	50.6%

TABLE 4: Four linked instances.

ID	Coexisting labels
1	Japan, Love, Animation, Magic, Youth, <i>Suzumiya Haruhi</i> , <i>Higashino Keigo</i> , <i>Makoto Shinkai</i> , <i>Da Vinci Code</i> , <i>Harry Potter</i>
2	American, Japan, Love, Classic, China, <i>Perfume</i> , <i>Pride and Prejudice</i> , <i>Shunji Iwai</i> , <i>Mayday</i> , <i>Fairy Tale</i>
3	Japan, Childhood, British, Magic, <i>Dragon Ball</i> , <i>Saint Seiya</i> , <i>Slam Dunk</i> , <i>Doraemon</i> , <i>Harry Potter</i> , <i>Garfield</i>
4	American, Love, British, Humanity, Hong Kong, Science Fiction, China, Youth, Growth, Japan

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The work is supported by the National Natural Science Foundation of China (Grant nos. 61309007, U1636219) and the National Key Research and Development Program of China (Grant no. 2016YFB0801303). The dataset is provided by Microsoft Research Asia.

References

[1] Facebook Inc, “Facebook quarterly report [sections 13 or 15(d)],” <https://www.sec.gov/Archives/edgar/data/1326801/000132680116000087/0001326801-16-000087-index.htm>.

[2] Twitter, Twitter quarterly report [sections 13 or 15(d)], <https://www.sec.gov/Archives/edgar/data/1418091/000156459016026749/000156459016026749/0001564590-16-026749-index.htm>.

[3] I. Inc, “600 Million and counting”.

[4] T. Ma, J. Zhou, M. Tang et al., “Social network and tag sources based augmenting collaborative recommender system,” *IEICE Transactions on Information and Systems*, vol. E98-D, no. 4, pp. 902–910, 2015.

[5] R. Zafarani and H. Liu, “Connecting corresponding identities across communities,” in *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media (ICWSM ’09)*, pp. 354–357, San Jose, Calif, USA, May 2009.

[6] R. Zafarani and H. Liu, “Connecting users across social media sites: a behavioral-modeling approach,” in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 41–49, ACM, Chicago, Ill, USA, August 2013.

[7] J. Vosecky, D. Hong, and V. Y. Shen, “User identification across multiple social networks,” in *Proceedings of the 1st International Conference on Networked Digital Technologies (NDT ’09)*, pp. 360–365, IEEE, Ostrava, Czech Republic, July 2009.

[8] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischofi, “Identifying users across social tagging systems,” in *Proceedings of the 5th Annual Conference on Weblogs and Social Media (ICWSM ’11)*, Barcelona, Spain, July 2011.

[9] F. Iqbal, H. Binsalleeh, B. C. M. Fung, and M. Debbabi, “Mining writeprints from anonymous e-mails for forensic investigation,” *Digital Investigation*, vol. 7, no. 1-2, pp. 56–64, 2010.

[10] Y. Zhong, N. J. Yuan, W. Zhong, F. Zhang, and X. Xie, “You are where you go: inferring demographic attributes from location check-ins,” in *Proceedings of the 8th ACM International Conference on Web Search and Data Mining (WSDM ’15)*, pp. 295–304, ACM, Shanghai, China, February 2015.

[11] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in *Proceedings of the IEEE Symposium on Security and Privacy (SP ’08)*, pp. 111–125, IEEE, Oakland, Calif, USA, May 2008.

- [12] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *Proceedings of the 30th IEEE Symposium on Security and Privacy*, pp. 173–187, IEEE, Berkeley, Calif, USA, May 2009.
- [13] H. Li, F. Tian, W. Chen, T. Qin, and T.-Y. Liu, "Generalization analysis for game-theoretic machine learning," <https://arxiv.org/abs/1410.3341>.
- [14] F. Tian, H. Li, W. Chen, T. Qin, E. Chen, and T.-Y. Liu, "Agent behavior prediction and its generalization analysis," <https://arxiv.org/abs/1404.4960>.
- [15] D. He, W. Chen, L. Wang, and T.-Y. Liu, "A game-theoretic machine learning approach for revenue maximization in sponsored search," <https://arxiv.org/abs/1406.0728>.
- [16] X. Sun, Y. Liu, J. Li, J. Zhu, H. Chen, and X. Liu, "Feature evaluation and selection with cooperative game theory," *Pattern Recognition*, vol. 45, no. 8, pp. 2992–3002, 2012.
- [17] N. J. Yuan, F. Zhang, D. Lian, K. Zheng, S. Yu, and X. Xie, "We know how you live: exploring the spectrum of urban lifestyles," in *Proceedings of the 1st ACM Conference on Online Social Networks (COSN '13)*, pp. 3–14, ACM, Boston, Mass, USA, 2013.
- [18] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.
- [19] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.
- [20] B. Gu and V. S. Sheng, "A robust regularization path algorithm for ν -support vector classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2016.
- [21] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 7, pp. 1403–1416, 2015.
- [22] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
- [23] Y. Chen and O. Kesten, "From boston to shanghai to deferred acceptance: theory and experiments on a family of school choice mechanisms in," in *Proceedings of the International Conference on Auctions, Market Mechanisms and Their Applications*, pp. 58–59, Springer, New York, NY, USA, April 2011.
- [24] P. Guillen and O. Kesten, "On-campus housing: theory vs. experiment."
- [25] A. E. Roth and E. Peranson, "The effects of the change in the NRMP matching algorithm," *The Journal of the American Medical Association*, vol. 278, no. 9, pp. 729–732, 1997.
- [26] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article no. 27, 2011.
- [27] X. Kong, J. Zhang, and P. S. Yu, "Inferring anchor links across multiple heterogeneous social networks," in *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, pp. 179–188, ACM, San Francisco, Calif, USA, 2013.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

