

MLP-IA: Multi-Label User Profile Based on Implicit Association Labels

Lingwei Wei¹², Wei Zhou¹²³, Jie Wen¹², Meng Lin¹², Jizhong Han¹² and Songlin Hu¹²

¹ School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

² Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

³zhouwei@iie.ac.cn

Abstract. Multi-Label user profile is widely used and have made great contributions in the field of recommendation systems, personalized searches, etc. Current researches on multi-label user profile either ignore the associations among labels or only consider the explicit associations among them, which are not sufficient to take full advantage of the internal associations. In this paper, a new insight is presented to mine the internal correlation among implicit association labels. To take advantage of this insight, a multi-label propagation method with implicit associations (MLP-IA) is proposed to get user profile. A probability matrix is first designed to record the implicit associations and then combine the multi-label propagation method with this probability matrix to get more accurate user profile. Finally, this method proves to be convergent and faster than traditional label propagation algorithm. Experiments on six real-world datasets in Weibo show that, compared with state-of-the-art methods, our approach can accelerate the convergence and its performance is significantly better than the previous ones.

Keywords: Implicit Association Labels, User Profile, Multi-Label Propagation.

1 Introduction

As the consumer economy enters an era of “information overload” [1], a large number of personalized service platforms, including recommendation systems, begin to emerge to satisfy human’s more individualized demands. User profile, the fully understanding of users, is the basis of recommendation system [19] and exact-marketing [2][3]. As a result, it is essential to provide an accurate and effective method to get user profiles.

Recently, user profile is widely studied by label propagation methods inferring user interests [4][5][6], user authority[21] and social attributes[20]. In order to get more accurate and abundant user profile, many researches prefer to apply multiple labels to analyze user profile. Some researches assumed labels were independent [5], ignoring the associations among them. However, it is not consistent with reality and cannot mine hidden label features very well. To overcome this limitation, Glenn et al.[1] considered the associations among labels to get user profile and obtained better performance. The associations in these works are explicit, indicating that there are some clear and definable connections among labels. For example, “Photography” and “Camera” are related, so they are called explicit association. As a result, a large number of methods based on

explicit association were proposed and have achieved good performance [22][23]. However, the existing methods rarely consider the implicit association of labels, where there are internal but not direct connections among them.

In many real-world applications, the associations among labels are complex [26]. With our observation, in addition to the explicit associations, there still exists some associations among implicit labels due to various reasons, such as uncertainty [24] or privacy issues [25]. For example, “Travel” and “health” are not related in any cases. However, they have many inherent connections, because people who like travelling always pay more attention to their health. Utilizing the implicit associations among labels, it is beneficial to make user profile more accurate and comprehensive.

To take advantage of this insight, a multi-label propagation method with implicit label associations (MLP-IA) is proposed to get user profile. We first design a probability matrix to record the implicit associations and then combine this probability matrix to multi-label propagation method to get more accurate user profile. Finally, we prove that the method is convergent and faster than traditional label propagation algorithms. To sum up, the main contributions of this study can be summarized as follows:

- **Insight.** We present a novel insight about associations among implicit association labels. In social platforms, due to users’ social and living habits, there are still certain implicit associations among labels. At the same time, mining the associations is useful for the construction of user profile.

- **Method.** A multi-label propagation method with implicit label associations is proposed to get user profile. We first design a probability matrix to record the implicit associations and then combine the multi-label propagation method with this probability matrix to get more accurate user profile. Finally, we prove that the method is convergent and faster than traditional label propagation algorithm.

- **Evaluation.** We conduct experiments to evaluate our method on six real Weibo data sets of different sizes. The comparative experiments evaluate the accuracy and effectiveness of the proposed method. The results show our method can accelerate the convergence and the performance is significantly better than the previous methods.

The following chapters are organized as follows: In section II, related works are elaborated in details. The section III explores our insights about the implicit association labels, and section IV describes the details of the proposed method and its efficiency. In section V, experiments and results are described. Finally, conclusions and future work are drawn in section VI.

2 Related Works

The existing researches in user profile can be divided into two parts. One is to infer user’s unknown attributes based on the user’s own data by text-mining methods, and the other is to propagate labels by social-network structure.

2.1 Text-mining Methods

There are many text-mining methods to extract user profile. The user's own data generally contains rich semantic information, so the user profile problem can be regarded as a text analysis problem [7][8]. For user's interest profile, most researchers apply the topic model (LDA) to complete the keyword extraction on the blog, and then use TFIDF algorithm to select features [9][10][11]. However, text mining often has high complexity, and the extracted profiles are unstable because of the richness of semantics.

2.2 Social-network Structure

The method is to label the unknown users based on the known users' labels by propagation, and multi-label algorithms are widely applied. Zhang et al. used multi-label propagation algorithm to mine user interests, and discovered potential interests of users through social relationships [6]. Xie et al. proposed the speaker-listener mechanism to update the label [12]. Dong et al. considered inference of gender and age using a social network, which is feasible only when the set of attribute values is extremely restricted[13]. To approach this, Chakrabarti et al. [14] proposed a method called EDGEEXPLAIN and Besel et al. built interest profile by propagating activation functions, and proved the method was more suitable than the most advanced methods [15]. Ma et al. innovatively introduced label propagation to improve the accuracy of the semi-supervised learning algorithm[16].

Some scholars found there were links between labels. Recently, the explicit associations among labels have been taken into consideration, Glenn et al. [1] introduced the explicit association labels and the results showed that their method performed well. To the best of our knowledge, there are few researches on the associations among implicit association labels.

3 Priori Knowledge with Implicit Association Labels

As is observed, in many social platforms, because of hot spot events or other special reasons, there are certain associations among implicit association labels. For example, as is shown in Fig.1, the node represents a user in Weibo and a directed edge indicates the user's social relationship. As highlighted with orange labels, we find the majority of users who like Entertainment in Weibo like Health as well.

We analyze the statistical characteristics of user interest labels in Weibo by correlation analysis [27] and a higher value indicates that there are certain associations among implicit association labels. We have show the top five label pairs in Fig.2 and the value is the correlation score of interest labels.

The statistical results are explicable. For example, Fig.2 shows that there are some associations between Health and Tourism. In the real world, users who like healthy lifestyle tend to pay more attention to tourism information. Obviously, they can enrich their lives through tourism and develop a healthy life with a relaxing lifestyle.

Based on our observation, implicit associations exists among labels and the features can be fully utilized to build a better user profile. Therefore, we will introduce the priori

probability among implicit association labels to improve user profile model. The details will be illustrated in next section.

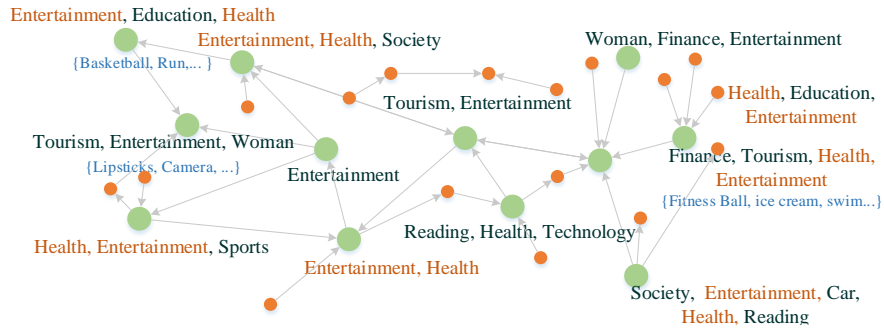


Fig. 1. An Example of interest labels propagation in Weibo. Note: nodes with labels indicate that users in Weibo where labels are the topic users are interested in. Green nodes are users with high influence such as V-plus users and orange nodes are some ordinary users. A directed edge indicates the user's social relationship. Some explicit labels are highlighted with blue color.

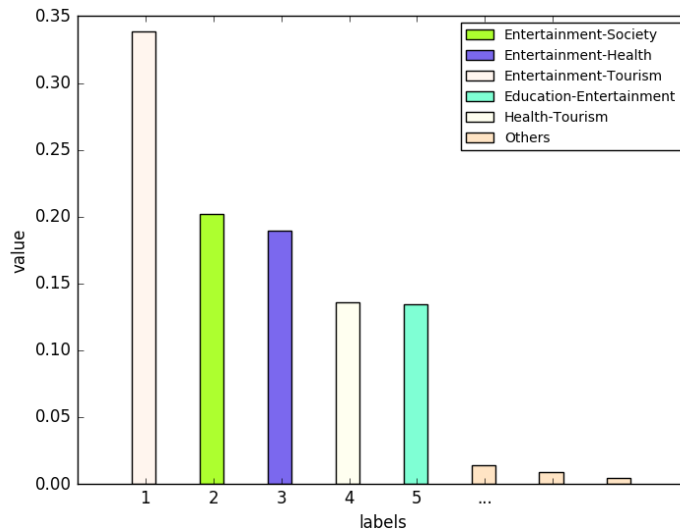


Fig. 2. Results of Labels Association Analysis.

4 Our Model

This section mainly focuses on the improved multi-label propagation. Firstly, we will construct the priori knowledge to introduce the associations among implicit labels. And

then two major matrixes in multi-label propagation algorithm will be initiation for propagation. Next, the model will be trained via labeled users and we will get the unlabeled users' label after a finite number of iterations.

The symbols mentioned in the paper are shown in Table.1.

Table 1. Symbols of our paper

Symbol	Description
I	A set of a pair of labels.
P	A matrix of priori probability of associations among implicit labels.
p_{ij}	An element of P , represents priori probability of implicit association between label i and label j .
$F^{(t)}$	The t -th iteration of <i>LABEL</i> . And the initialization is recorded as $F^{(0)}$.
T	A matrix of probability transferred among users.
R	A set of users' social relationships and the element (u_i, u_j) represents u_i follows u_j . $R = \{(u_i, u_j) u_i, u_j \in U\}$.
<i>FANS</i>	A set of users' fans and the element $FANS_i$ represents the fans set of user i .
<i>FOLLOW</i>	A set of users' followers and the element $FOLLOW_i$ represents the user set of user i follows.
C	A matrix that represents the connections between users.
c_{ij}	An element of C , represents whether there is a connection between user i and user j .

4.1 Priori Knowledge of Implicit Association Labels

From Section III, a new insight about associations among implicit association labels has been found. We analyze the interest labels of users in Weibo and find there is a certain connection among different interest labels.

Specifically, we define the priori probability matrix P as Eq.1 shows. The higher the value of p_{ij} is, the higher the probability of propagation among labels becomes.

$$p_{ij} = \frac{|\{t | t \in I \wedge (l_i, l_j) \subseteq t\}|}{Z} \quad (1)$$

where $Z = \sum_{i=0}^m \sum_{j=0}^m |\{t | t \in I \wedge (l_i, l_j) \subseteq t\}|$. Some scholars have proved that the associations in social network are complex due to various reasons, such as uncertainty [24] or special events[25]. Therefore, we define that elements of I by co-occurrence, cultural associations, event associations or custom associations and so on. The detail is shown in Eq.2.

$$I = I_1 \cup I_2 \cup I_3 \cup \dots \quad (2)$$

where I_i ($i = 1, 2, 3, \dots$) represents respectively a set of each user's interest label set, label sets sampled by cultural associations, event associations and custom associations.

4.2 Introduction of the Priori Knowledge

As our observation, in addition to other users' influence, the label will spread according to both other users and other labels, that is $F = U \cdot L \cdot P$. The labels will be propagated among users and each iteration is given by Eq.3, where λ is a hyper parameter and it controls the influence of initialization.

$$F^{(t+1)} = \lambda T \cdot F^{(t)} \cdot P + (1 - \lambda)F^{(0)} \quad (3)$$

The hyper parameter λ controls the influence of initialization. In each iteration, the users' labels will be updated by the neighbor node's labels and the implicitly associated labels. It's noted that after each iteration, the matrix F will be corrected by F_a for next correct propagation, which shows that our model is a semi-supervised learning method.

The loss function in the model uses the squared distance, as is shown in Eq.4.

$$\text{loss} = |F^{(t+1)} - F^{(t)}|^2 + \frac{1}{2} \|1 - \zeta\|^2 \quad (4)$$

In real world network, it is difficult to construct complete structure of networks because of privacy security. Therefore, we consider the influence of the integrity of social networks. In the model, a hyper parameter ζ is introduced.

We define $\zeta = \frac{\text{the number of relationships in dataset.}}{\text{the number of relationships in the real world.}}$, which represents the sparsity of social networks and the value indicates the integrity of a social network. When more relationships are added for constructing graph, ζ will tend to 1 and the model will get smaller loss value accordingly to make a better user profile.

4.3 Multi-Label User Profile Based on Implicit Association Labels

Traditionally, given labels set $L = \{l_1, \dots, l_m\}$, $U = \{u_1, \dots, u_a, \dots, u_{a+b}\}$, which contains a users with labels and b users without labels and their labels matrix $F = [F_a; F_b]$.

Firstly, if nodes are in a graph, multi-label propagation algorithm infers the labels via the aggregate labels of their neighbors until labels for all the nodes do not change [27]. The key to inference is the probability between nodes. In our model, two major matrixes, the user's initial interest vector matrix and label transfer matrix, are constructed. The details of two matrixes are as follows.

Initial Interest Vector Matrix. The initial interest vector matrix contains two parts. The one-hot method is used to build the labeled users' initial vector of interest labels. As is shown in Eq.5, if the user is with the label, the value will be 1. Otherwise, it will be 0.

$$f_{ij} = \begin{cases} 1 & \text{if the user } i \text{ is with the label } j \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The unlabeled users' initial vector is zero vector. From this, we transform the multiple interest labels into the multi-label vector for propagation. It is worth noting that

after each update iteration of label propagation algorithm, it is necessary to correct the labeled users' interest label matrix so as to obtain more accurate results of propagation.

Probability Transfer Matrix. The interest labels will spread among users. The label transfer matrix T is constructed via social relationships shown in Eq.6. Here we introduce two matrixes D and W for better convergence, which had proved in [17]. Elements of D are computed by Eq.7 and elements of W are computed by Eq.8.

$$T = D^{-1/2}WD^{-1/2} \quad (6)$$

$$d_{ii} = \sum_{j=1, \dots, n} w_{ij} \quad (7)$$

$$w_{ij} = \begin{cases} c_{ij} \times \text{Sim}(i, j) & i \neq j \\ 0 & i = j \end{cases} \quad (8)$$

where $\text{Sim}(i, j)$ indicates the similarity between user i and user j is expressed as is shown in the Eq.9. That is, the less the ratio of the value is, the closer the distance is.

$$\text{Sim}(i, j) = \frac{1}{\left| \log\left(\frac{|FANS_i|}{|FOLLOW_i|}\right) - \log\left(\frac{|FANS_j|}{|FOLLOW_j|}\right) \right| + 1} \quad (9)$$

Next, we will train the model via labeled users and the loss function is defined in Eq.4. After each iteration in Eq.3, the model will stop until loss is less than threshold that we set. The method is proved to be convergent in 4.4. The specific algorithm flow is shown in Algorithm 1.

Algorithm 1 MLP-IA

Require: $L = \{l_1, \dots, l_m\}$, $U = \{u_1, \dots, u_a, \dots, u_{a+b}\}$, $I_i (i = 1, 2, 3, \dots)$

- 1 Construct user interest label matrix F via Eq.5
- 2 Estimate the integrity of social networks ζ
- 3 Construct set $\{t | t \in I \wedge (l_i, l_j) \subseteq t\}$
- 4 Construct priori probability matrix P via Eq.1
- 5 Estimate $\text{Sim}(i, j)$ based on data sets.
- 6 Construct the matrix D and W via Eq.7 and Eq.8
- 7 Learn the probability transfer matrix T via Eq.6
- 8 **while** $\text{loss} > \text{THRESHOLD}$ **do**
- 9 $F^{(t+1)} = \lambda T \cdot F^{(t)} \cdot P + (1 - \lambda)F^{(0)}$
- 10 Fixed Source User Interest Label F_a
- 11 Update loss via Eq.4
- 12 **end while**
- 13 Output F_b

4.4 Analysis of Algorithms

Convergence analysis. The convergence of our method is shown as follows. Let the user's label matrix be F . According to the labeled and unlabeled users, T can be divided

into sub-matrices as is shown in Eq.10, where subscript ‘‘a’’ indicates the user’s label is known and subscript ‘‘b’’ indicates the user’s label is unknown.

$$\mathbb{T} = \begin{bmatrix} T_{aa} & T_{ab} \\ T_{ba} & T_{bb} \end{bmatrix} \quad (10)$$

From the section above, we can see that the core formula of the label propagation algorithm proposed in this paper is Eq.3. P matrix represents the co-occurrence relationship of labels, and $0 \leq p_{ij} \leq 1$. F_a is the interest label of the source user, which is fixed and invariant. Therefore, our method is simplified as $F_b \leftarrow T_{bb}F_bP + T_{ba}F_aP$, which leads to Eq.11.

$$F_b = \lim_{n \rightarrow \infty} (T_{bb})^n F_b^{(0)} P^n + \left(\sum_{i=1}^n (T_{bb})^{(i-1)} \right) T_{ba} F_a P^n \quad (11)$$

where $F_b^{(0)}$ is the initial value of F_b . Because T matrix is row-regular, T_{bb} is a submatrix of T , so it follows Eq.12. Therefore, we can get Eq.13.

$$\exists \gamma < 1, \sum_{j=1}^u (T_{bb})_{ij} \leq \gamma, \forall i = 1, \dots, u \quad (12)$$

$$\begin{aligned} \sum_j (T_{bb})_{ij}^n &= \sum_j \sum_k (T_{bb})^{(n-1)}_{ik} (T_{bb})_{kj} = \sum_k (T_{bb})^{(n-1)}_{ik} \sum_j (T_{bb})_{kj} \\ &\leq \sum_k (T_{bb})^{(n-1)}_{ik} \gamma \leq \gamma^n \end{aligned} \quad (13)$$

P is priori knowledge and $0 \leq p_{ij} \leq 1$, which can accelerate convergence. And $(T_{bb})^n$ indicates the sum of each line converges to 0, from which we can conclude in $(T_{bb})^n F_b^{(0)} P^n \rightarrow 0$. Thus the initial value of $F_b^{(0)}$ is inconsequential. Obviously, $F_b = (1 - T_{bb})^{-1} T_{ba} F_a$ is a fixed point.

Time complexity analysis. In the initialization of the label propagation algorithm, we need to establish an initial label for each Weibo user, and the complexity of the process is $O(n)$. In the propagation of the interest label, due to the convergence of our method, the iterations are fixed. And the time complexity of each iteration is $O(m)$. Therefore, the entire algorithm is nearly linear in time complexity.

5 Experiments

5.1 Dataset

Like Twitter, Weibo is the largest social network platform in China. To prove the universality and effectiveness of our method, we evaluate our method in different scale data sets in Weibo.

Firstly, we randomly get six different sets of users and their social datas such as followers, fans and blogs etc. The scale of the data sets are illustrated in Table 2. And

different data sets are collected at different times. Due to the limit of Weibo, we just get part of their following users and obtain sparse social relationships of users.

Table 2. Dataset of our paper

Dataset	Number of Labeled Users	Number of Unlabeled Users	Size
1#	1129	4001	5130
2#	1355	4801	6156
3#	1581	5601	7182
4#	1807	6401	8208
5#	2033	7201	9234
6#	2200	8700	10900

Table 3. Baselines of our paper

Notation	Description
MLP	Multi-label propagation with any association among labels
MLP-EA	Introduce associations among explicit association labels
MLP-EIA	Introduce associations among both explicit and implicit association labels

Next, according to the characteristics of Weibo, we artificially labeled users' interest with interest labels based on their blogs and social relationships. The labeled users are selected if the user is marked with a "V" which means his identity had been verified by Sina. Analyzed by Jing et al.[18], these users were very critical in the propagation.

5.2 Comparisons and Evaluation Setting

To evaluate the performance of our method (MLP-IA), we compare it with other methods. Table 3 lists some compared baselines. We first compare with traditional Multi-Label Propagation(MLP) to evaluate the effectiveness of priori knowledge. Then we select Multi-Label Propagation Based on Explicit Association Labels (MLP-EA) to evaluate whether implicit association labels can perform better to mine the relationship than explicit association labels learning from [1]. In baseline three: Multi-Label Propagation Based on Explicit and Implicit Association Labels (MLP-EIA), we introduce associations among both explicit and implicit association labels to experiment, fully exploring the relationships among labels. Finally, to explore the impact of social relationship integrity on results, we make some new experiments in 5.4.

In the experiments, we will analyze the precision ratio and recall ratio of method which respectively represent the accuracy and comprehensiveness of user profile. And F1-Measure is a harmonic average of precision ratio and recall ratio, and it reviews the performance of the method.

5.3 Results and Analysis.

The experiment results are shown in Fig.3. From Fig.3, we can see MLP has the highest precision ratio, and MLP, MLP-IA and MLP-EIA have stable result. However, as the size of data set increases, the precision ratio of MLP-EA decreases continuously. In the

recall ratio, MLP-IA and MLP-EIA perform better than others and MLP have the worst result. Furthermore, in the F1-Measure, MLP-IA and MLP-EIA, introducing priori knowledge of association labels, achieve the best results.

Compared with MLP, the results show that although the precision ratio of our method is slightly reduced due to the introduction of priori knowledge, the recall ratio has been greatly improved, and the F1-Measure has also been improved. Therefore, the results prove that the association among labels is effectively mined based on implicit association labels and the user’s interests can be well mined. But the recall ratio is less than MLP-IA. It indicates that although MLP can predict the interest label accurately, user profile is not complete and the convergence speed of model is rather slow.

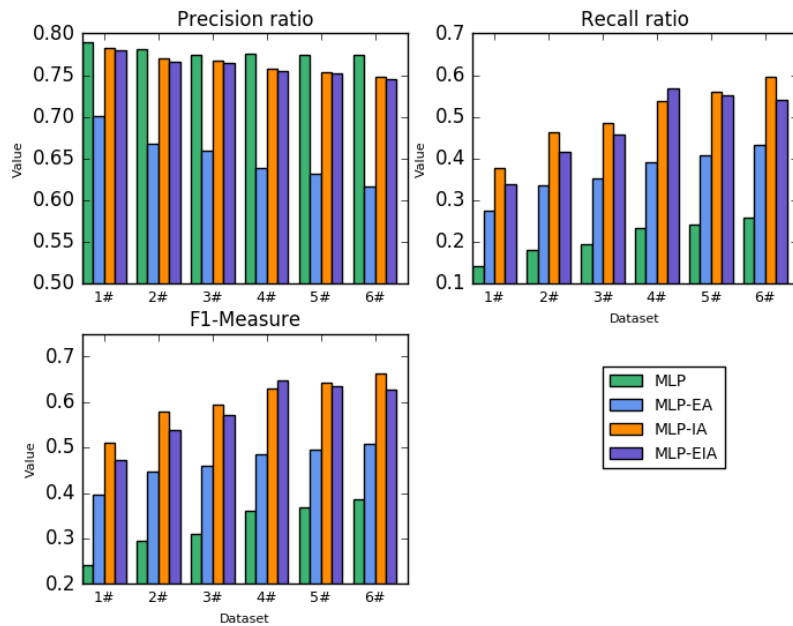


Fig. 3. Results of the precision ratio, the recall ratio and the F1-Measure.

In the MLP-EA, the associations among explicit labels was considered. The associations were calculated by word2vec before the application of label propagation. Results show that with the introduction of explicit associations among labels, the precision ratio basically remains unchanged and the recall ratio is improved especially in larger data sets, which proves the effectiveness of the explicit mining.

However, as the results show, our method performs better in recall ratio and F1-Measure. It can indicate association among implicit labels can perform better to mine the relationships among labels than explicit labels. In Weibo, posts are more arbitrary and it provides more features to make user profile. Nevertheless, there is much “noise” disturbing the results. Considering priori knowledge of association among implicit labels can avoid introducing textual “noise”. On the other hand, the semantics of posts are too diverse to mine the associations among implicit association labels well. Instead

of considering too many explicit details, it is more beneficial to explore the associations among implicit association labels.

Furthermore, we consider both the explicit labels and implicit association labels in MLP-EIA. The results show that our method has similar performance with MLP-EIA. It can prove that our method MLP-IA, introducing the implicit association labels, can capture the feature of users deeply and make user profile well. As is mentioned above, the associations among explicit labels includes too many features which are positive or negative for model training. So the model could not identify them well. Therefore, user profile model can be well constructed only by introducing implicit associations among labels.

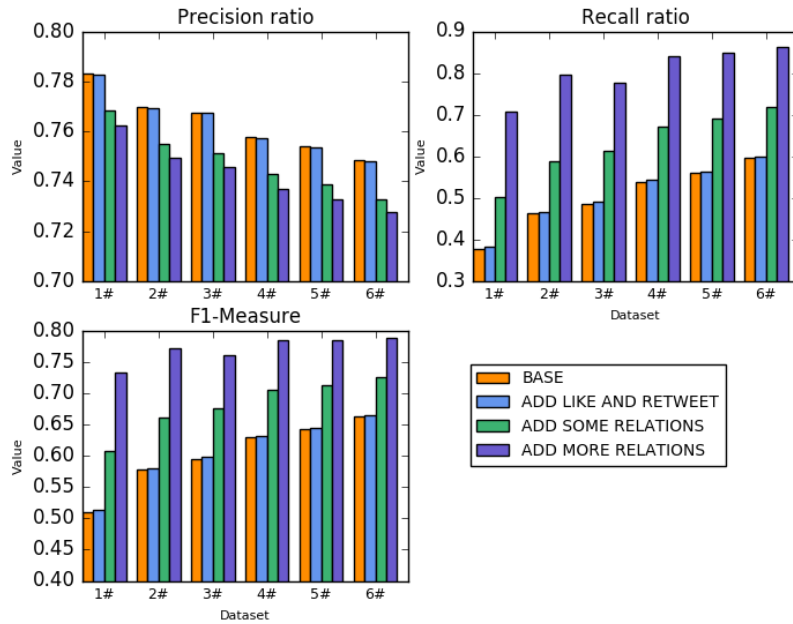


Fig. 4. Further Results: the influence of social relationship integrity.

5.4 The Influence of the Social Relationship Integrity

To explore the impact of social relationship integrity on results, we conduct new experiments by adding more social relationships based on the same users' sets. In addition, we consider to add relationships of "LIKE" and "RETWEET", which also represents the interaction among users. The results are shown in Fig.4. We can see that with the increase of known social relationship data, the performance of our method is gradually improving, especially in recall ratio. It proves that the model can identify more interest labels according to social relationship data.

The hyper parameter ζ in Eq.7 which represents the sparsity of social networks. When adding more social relationships, the model will work on the more complete social network graph. The param will be equal to 1 and lead to a smaller loss. On the

other hand, the matrix T in Eq.6 will be dense after adding more social relationships, and the model can capture more features among users after each iteration.

Therefore, based on more relationships among users, we can build a more complete social network graph accordingly and explore more information about these interaction. And the richer the social relationships is, the higher the recall ratio of interest profile is.

5.5 Convergence

To evaluate the performances between the introduction of explicit labels and implicit association labels, we experiment with the convergence of the iterative times and time consuming. And the results are shown in Fig.5.

The results intuitively show that our method converges faster with the introduction of the priori knowledge. And as the data scale increases, the time complexity is still low and time consumption does not increase exponentially. In particular, by comparison, our method takes the least iteration time. Therefore, the introduction of associations among implicit association labels can accelerate the result convergence, and accordingly it can be used in more real-time and large-scale recommendation systems.

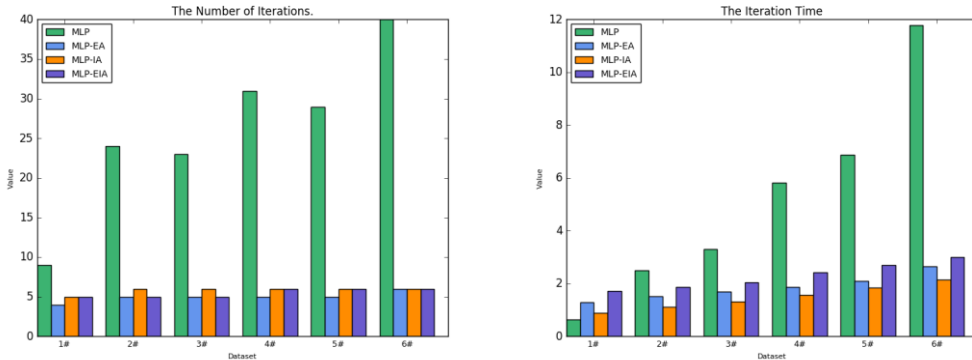


Fig. 5. The Number of Iterations and The Iteration Time. The left is the number of iterations and the right is the iteration time.

5.6 Summary

We compare our method with three baselines. The results show that our method can accelerate the convergence of propagation and make a significant increase in F1-Measure. However, baseline MLP-EA does not perform better in the experiments. The reason is, in the real world, explicit associations among labels include too many features which may be positive or negative for model training. The model could not identify them well. Instead of considering too many explicit details, it is more beneficial to explore the associations among implicit association labels.

Furthermore, our method achieves a similar performance with MLP-EIA. However, MLP-EIA takes much more time than our method. It proves that, although explicit

associations among labels contains more features, it did not work in our model. Therefore, user profile could be well constructed by only introducing implicit associations among labels.

In addition, we explore the impact of social relationship integrity on results. The results show that, with the increase of known social relationship data, the performance of our method is gradually improving and it can identify more interest labels according to social relationship data.

6 Conclusion and Future Work

In this paper, we have studied the user profile by multi-label propagation. We proposed an improved multi-label propagation algorithm to utilize implicit association among labels and the implicit association labels can demonstrate more relationships among users. The experiments based on six real-world Weibo datasets have shown that our method accelerates the convergence and gets better performance than the previous methods.

Future work will pay more attention to improve the recall ratio of our method by extending the social relationships.

References

1. Boudaer, G., Loeckx, J.: Enriching Topic Modelling with Users' Histories for Improving Tag Prediction in Q&A Systems. In: 25th International Conference Companion on World Wide Web, pp. 669–672. ACM, USA (2016).
2. Paulo, R., S., Frederico, A., D.: RecTwitter: A Semantic-Based Recommender System for Twitter Users. In Proceedings of the 24th Brazilian Symposium on Multimedia and the Web. ACM, USA(2018).
3. Nurbakova, D.: Recommendation of Activity Sequences during Distributed Events. In: Conference, pp.261-264 (2018).
4. Chang, P., S., Ting, I., H., Wang, S. L.: Towards Social Recommendation System Based on the Data from Microblogs. In: International Conference on Advances in Social Networks Analysis and Mining, pp.672-677. IEEE. USA (2011).
5. Li, R., Wang, C., Chang, C. C.: User profiling in an ego network:co-profiling attributes and relationships. International Conference on World Wide Web, pp.819-830. ACM, USA (2014).
6. Zhang, J., L.: Application of tag propagation algorithm in the interest map of weibousers. Programmer (5), 102-105 (2014).
7. Yao, Y., Yao, T.: Gender Classification of Chinese Weibo Users. In: Proceedings of the 2017 International Conference on E-commerce, E-Business and E-Government, pp. 5-8. ACM, USA (2017).
8. Xu, Y., N.: Modeling and application of social network user portrait based on text mining. Beijing University of Posts and Telecommunications, China (2016).
9. Chen, Z., M., Hu, Z., Y.: User Profile Study on UGC Website. Computer Systems & Applications (2017).
10. Qin, F., Chen, Z., Zheng, X. User Behavior Prediction Based on Emotion and Interest. Computer System Application 27(1), 28-34(2018).

11. Tu, K., Ribeiro, B., Jensen, D., et al.: Online dating recommendations: matching markets and learning preferences. In: Proceedings of the 23rd International Conference on World Wide Web, pp.787-792. ACM, USA(2014).
12. Xie, J., Kelley, S., Szymanski, B. K. Overlapping community detection in networks:the state-of-the-art and comparative study. *ACM Computing Surveys* 45(4), 1-35 (2011).
13. Dong, Y., Yang, Y., Tang, J., et al.: Inferring user demographics and social strategies in mobile social networks. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.15-24. ACM, USA(2014).
14. Chakrabarti, D., Funiak, S., et al.: Joint Label Inference in Networks. In: Companion of the the Web Conference, pp.483-487 (2017).
15. Besel, C., Granitzer, M. On the quality of semantic interest profiles for online social network consumers. ACM, USA (2016).
16. Ma, C.: A method of social network user portrait analysis based on topic model. Anhui, China university of science and technology, China (2017).
17. Zoidi, O., Fotiadou, E., Nikolaidis, N., et al. Graph-based label propagation in digital media: a review. *Acm Computing Surveys*, 47(3), 48 (2015).
18. Jing, M., Yang, X., X.: The Characterization and Composition Analysis of Weibo "V". *News and Writing* 2, 36-39 (2014).
19. Fernando, A., Ashok, C., Tony, J., et al. Artwork Personalization at Netflix. In Proceedings of the 12th ACM Conference on Recommender Systems, pp487-488. ACM, USA (2018).
20. Chakrabarti, D. , Funiak, S. , Chang, J. , et al. Joint inference of multiple label types in large networks. Eprint Arxiv, 874-882. (2014).
21. Pal, A. , Herdagdelen, Amaç, Chatterji, S. , et al. Discovery of Topical Authorities in Instagram. International Conference on World Wide Web International World Wide Web Conferences Steering Committee, (2016).
22. Liang, S., S., Zhang, X., L., et al. Dynamic Embeddings for User Profiling in Twitter. In Proceedings of the 24th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2018), pp. 1764-1773, UK, (2018).
23. Roberto, P., S., and Frederico, A., D.. RecTwitter: A Semantic-Based Recommender System for Twitter Users. In Proceedings of the 24th Brazilian Symposium on Multimedia and the Web (WebMedia '18). pp371-378, ACM, USA, (2018).
24. Peng Peng, Raymond Chi-Wing Wong, and Phillip S Yu. Learning on probabilistic labels. In Proceedings of the 2014 SIAM International Conference on Data Mining. SIAM, pp307–315.(2014).
25. Arun Shankar Iyer, J Saketha Nath, and Sunita Sarawagi. Privacy-preserving class ratio estimation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp925–934, (2016).
26. Huai, M., Miao, C., Li,Y., et al. Metric Learning from Probabilistic Labels. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18). ACM, USA, pp1541-1550. (2018).
27. Zhu, X., Ghanramani Z. Learning from labeled and unlabeled data with label propagation [C]. Pittsburghers: Carnegie Mellon University, (2002).