NLPCC 2018 Shared Task User Profiling and Recommendation Method Summary by DUTIR_9148

Xiaoyu Chen, Jian Wang(✉), Yuqi Ren, Tong Liu, and Hongfei Lin

Dalian University of Technology, Dalian Liaoning 116023, China
wangjian@dlut.edu.cn

Abstract. User profiling and personalized recommendation plays an important role in many business applications such as precision marketing and targeting advertisement. Since user data is heterogeneous, leveraging the heterogeneous information for user profiling and personalized recommendation is still a challenge. In this paper, we propose effective methods to solve two subtasks working in user profiling and recommendation. Subtask one is to predict users' tags, we treat this subtask as a binary classification task, we combine users' profile vector and social Large-scale Information Network Embedding (LINE) vector as user features, and use tag information as tag features, then apply a deep learning approach to predict which tags are related to a user. Subtask two is to predict the users a user would like to follow in the future. We adopt social-based collaborative filtering (CF) to solve this task. Our results achieve second place in both subtasks.

Keywords: User tags prediction · User following recommendation User modeling · Collaborative filtering · Deep learning

1 Introduction

With the rapid development of the Internet, a large amount of user information has been generated on the Internet. User profiling [1] can effectively use this information to analyze user's attributes, and it is widely used both in recommendation system and precise advertisement [2]. Social media is one of the most important components of Internet, and it has become essential for people's life. People like to share all kinds of information through social applications, which makes the user research based on social media become more and more important. Users usually use a set of phrases called tag to indicate their personal characteristics such as hobbies and interests and characters on social network. Many social networking services provide tag recommendation function for users to help them set the tags. We can use users' tags to solve problems such as community discovery [3]. Meanwhile, users like to make friends who have a lot in common through social networks. And friend recommendation is a very important function of social networking services, therefore, it has been the focus of social network research.

M. Zhang et al. (Eds.): NLPCC 2018, LNAI 11109, pp. 420–428, 2018.
https://doi.org/10.1007/978-3-319-99501-4_38

This shared task contains two subtasks. Subtask one gives users' other information except tags, we need predict which tags are related to a user. Subtask two gives users' following relationship and other provided information, we need predict the users a user would like to follow in the future. We regard subtask one as a classification problem and subtask two as a recommendation problem. We will introduce the two subtasks respectively in the following sections.

1.1 Subtask One

As an important function of social networking services, user tag recommendation has become the focus of research. The classic methods include content-based user tag recommendation method [4, 5] and graph-based user tag recommendation method [6–8]. Yamaguchi et al. [9] proposed a method to predict the tags which are related to a user by using Twitter list. Twitter list is an official functionality to list users may share information about the topic represented in list name. So they extracted the phrases from list names as user's tags, and exploited the relationship between tags and users, then recommended the most relevant tags to the users. Wu et al. [10] considered all of user's tweets as a document, then adopted TextRank [11] to extract keyphrases from it and treated the keyphrases as user's tags. In addition to tweets, there are complex relationships and links among users in social networks, which can be used in user tag recommendation. Lappas et al. [12] believed that users had many different interests. The reason why a user follow others is that they have the same interests. Therefore, they used the underlying social endorsement network to extract useful tags for users. And Chen et al. [13] proposed a method that recommended tags of the users who a user followed to a user. Because they considered that a user tended to follow the users who had the same interests with him. Since deep learning has been successfully used in many research fields, it is gradually applied to user tag recommendation task.

For subtask one, we propose a deep learning approach to predict which tags are related to a user. There are three parts in our approach, we will elaborate in the next section.

1.2 Subtask Two

Friend recommendation is a very popular personalized service in social networking services. Collaborative filtering(CF) [14] is the classic approach to the recommendation problem, which is comprised of two main methods, model-based CF and memory-based CF. The latter includes user-based CF [15] and item-based CF [16]. And friend recommendation is also a recommendation problem, there is some research on it. Bian et al. [17] proposed a method called MatchMaker, which was based on personality matching and collaborative filtering. It leveraged the mutual understanding and social information among users in social networks. Agarwal et al. [18] proposed an implicit rating model, for estimating a user's affinity toward his friends, which uncovered the strength of relationship, utilizing both attribute similarity and user interaction intensity. It was also a CF-based framework. Users usually share their check-in location in social networks, and the check-in data can be used for friend recommendation. Lin et al. [19] considered that customary location a user checked in can represent his social circle, so they achieved friend recommendation based on social circle similarity which was computed by the distance between the two geographical locations.

For subtask two, we propose a social-based collaborative filtering method and rules to realize users' friend recommendation. We will elaborate in the next section.

2 Method

2.1 Subtask One

We regard this subtask as a binary classification task and propose a deep learning framework to solve it. We encode user's personal information such as age, province and city into a vector, use the Large-scale Information Network Embedding(LINE) [20] to encode the social network information into a vector, and join these two vectors into a vector as the user features. Although a user's tags are desensitized, they still contain a lot of semantic information. By encoding them into a vector as the tag features, and then we join these two features as the sample feature, finally we apply our model to achieve the subtask.

As is shown in Fig. 1, the architecture of our model consists of three components: (1) Embedding layer to encode user and tag information; (2) Fully connected layer to obtain user and tag features from the embedding layer; (3) Output layer to predict a user's tags.

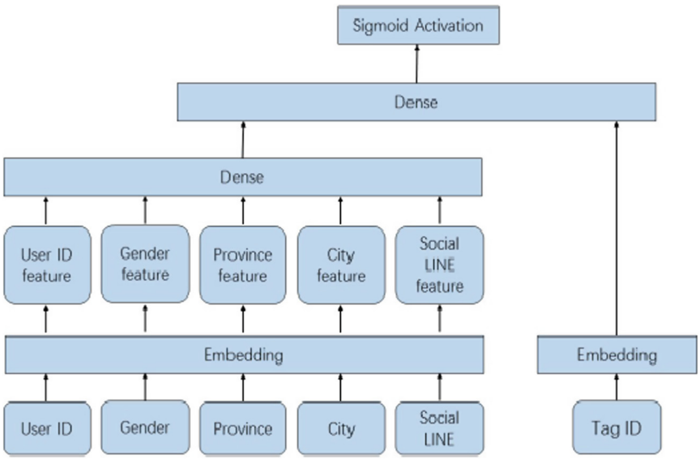

Fig. 1. Architecture of our proposed method

Embedding Layer

Embedding approach is convenient for us to capture the relationship between key information from data. Therefore, each information about user and tag is encoded into a real-valued vector by build embedding matrix $I^{emb} \in \sum_{i}^{n} V^{i} \times N$ in the embedding layer, where V^i denotes the dimension of the each information vector and N is the number of users.

We mainly adopt profiles data as the user information to generate user embedding vector, including user ID, gender, province and city. According to the data of the social ties, one user has followed many users. We connect users who have social relationships with each other and eventually form a social network. We use LINE approach to map all the users nodes in the network to a d-dimensional vector, and try to keep the original network structure. Finally, we merge the user IDs, genders, provinces, cities and social information embedding vectors together as user matrix.

For tag information, tag ID can reflect the relevant characteristics of tag. So the tag matrix consists of the embedding vector of tag IDs.

Fully Connected Layer
We extract user features from the merged user matrix by fully connected layer. Next, we merge the user features with the tag features. We use the fully connected layer again to get connection between user and tag.

Output Layer
Finally we use sigmoid function to predict a user's tags. The sigmoid function is a probabilistic classification algorithm that classifies by probability distribution. The calculation result is from 0 to 1. We consider that if the output probability is greater than 0.5, the tag is related to the user, otherwise the tag is not related to the user.

2.2 Subtask Two

Friend recommendation based on the proportion of common friends
We usually know that, if A and B are friends, and B and C are friends, then A is also likely to be willing to make friends with C. Therefore, we can achieve friend recommendation by calculating the proportion of common friends of two users. But there is also a problem here, the ratio of the total number of friends of a user to the number of common friends is usually very likely to have a certain influence on the friends to be recommended. So first, we use the ratio of the number of common friends to the total number of friends to weight them, and then recommend top-10 of results to a user.

Friend recommendation based on social-based collaborative filtering
As the saying goes, birds of a feather flock together, and people always like to make friends with like-minded people. The like-minded people mean that the people who have interests in the same. Therefore, we can also recommend friends by computing the similarity between users.
We use collaborative filtering (CF) to achieve the similarity computation between uses. One of the main functions of collaborative filtering is to make recommendations. Users can be analyzed based on their historical data. And recommend similar users based on their different preferences. Collaborative filtering is divided into user-based collaborative filtering algorithms and article-based collaborative filtering algorithms.

The key to achieve the friend recommendation through a social-based CF method is to find users similar to the target user. In this subtask, we use the data which describes user's following relationship to compute similarity between users. We denote the user set as $U = \{u_1, u_2, u_3, \ldots, u_m\}$, and denote every user's following user set as $F = \{f_1, f_2, f_3, \ldots, f_n\}$, at the same time, we also create a matrix $S \in R^{M \times N}$ indicates users and the users they follow, and if user i follows user j, $S_{ij} = 1$, otherwise $S_{ij} = 0$. To simplify the calculations, we normalize each line of the matrix S to get a new similarity matrix M_{uf}.

$$M_{uf} = Norm(S) \cdot Norm(S)^T \tag{1}$$

Where,$M_{uf} \in R^{M \times M}$ and each element m_{ij} represents the similarity between user i and user j. With the user similarity, we can use another matrix multiplication to obtain the score of new friends for each user as follow:

$$P = M_{uf} \cdot S - S \tag{2}$$

Where, $P \in R^{M \times N}$ and each element p represents the score for user i and user j.

Friend recommendation based on the most-popular rule
On social networking platforms, such as Weibo, people tend to follow well-known users whose common feature is that they have a high number of followers, so we use this rule to make friend recommendation and recommend the most popular users to a user.

3 Experiments and Results Analysis

3.1 Subtask One

Dataset and Evaluation
We use profiles data, tags data and socials data for subtask one. The number of unique users in the tags file is 11995. We split tags data to form an offline training set and test set by users group. For all users, split out 75 percent of the users with tags from tags data to get training set, the rest as the offline test set. Since the provided data contains only positive samples, we choose the top 20 tags of the frequency of occurrence to build negative samples. If user has no relationship with the top 20 tags, the combination of the user and tag serves as a negative sample. Our offline training set contains 9,796 users, each user has multiple tags, the total number of combination of users and tags is 176,619, the number of negative samples is 117,746, the number of positive samples is 58,873. The rest 2,399 users with tags for the test set, the total number of data is 47,980. The number of training set tag space is 18,496, and the tag space of test set is top 20.

The dimension of user ID's embedding vector is 128. We map the gender f and m to 1 and 0 respectively. The dimensions of embedding vector of province and city are both 32. Each node in the user social network represents a user. We use the LINE method to train the node vector and set the vector dimension to 100.

The quality of subtask one is evaluated by $F1@K$ (K = 3), and we also list $P@K$, $R@K$ for analysis. The calculation formula is as follows:

$$P_i@K = \frac{|H_i|}{K} \tag{3}$$

$$R_i@K = \frac{|H_i|}{|V_i|} \tag{4}$$

$$F1_i@K = \frac{P_i@K \times R_i@K}{P_i@K + R_i@K} \tag{5}$$

$$\mathrm{P}@K = \frac{1}{N}\sum\nolimits_{i=1}^{N} P_i@K \tag{6}$$

$$\mathrm{R}@K = \frac{1}{N}\sum\nolimits_{i=1}^{N} R_i@K \tag{7}$$

$$\mathrm{F1}@K = \frac{1}{N}\sum\nolimits_{i=1}^{N} F1_i@K \tag{8}$$

Where $|H_i|$ is the correctly predicted tags for user i's top K prediction, $|V_i|$ is the correct tags for user i. $P_i@K$, $R_i@K$, $F1_i@K$ is the precision, recall and F1 for a user i. N is the user count.

Experimental Results and Analysis
We evaluate our method in test set. First we evaluate the deep recommendation model with user features from profiles data. Then we merge the features of the social network with the trained user features through fully connected layer. Finally, we merge user's social feature and profile information as user features together. The results are shown in Table 1.

Table 1. The result of each method (K = 3).

Method	P@K	R@K	F1@K
Profile	4.175%	3.349%	3.652%
Fully connected layer + social	5.669%	3.418%	3.876%
profile + social	5.992%	3.505%	4.037%

Based on the results in Table 1, we find that the method combine profile data with user social information as embedding layer has the best performance. This demonstrates that social network information is effective in tag recommendation. The social information can better represent the commonalities between users.

3.2 Subtask Two

Dataset and Evaluation
We use social network data for subtask two. We split social data to form an offline training set and test set by friend(user2) groups. For each friend, split out 75 percent of

the social data to get the training set, so that all the friend already appear in the training set. Then we filter out the users who don't exist in the training set from the rest of social data to become test data set.
The quality of subtask two is evaluated by $F1@K$ (K = 10), and we also list $P@K$ and $R@K$ for analysis. The calculation formula is the same as subtask one.

Experimental Results and Analysis
We evaluate our method on the offline test set. The results are shown in Table 2.

Based on the results in Table 2, we find that the social-based CF has the best performance. This demonstrates that social network information is effective in User Following Recommendation. Because the users that have common friends are similar. It is effective to recommend the friends that the similar users have to a user.

Table 2. The result of each method (K = 10).

Method	P@K	R@K	F1@K
The proportion of common friends	0.202%	0.210%	0.213%
Social-based CF	0.679%	0.762%	0.679%
Most-popular rule	0.642%	0.747%	0.640%

4 Conclusion and Future Work

In this paper, we elaborate our methods and ideas on User Profiling and Recommendation shared tasks. For subtask one, we take it as a classification problem and adopted a deep learning method to predict user's tags. We focus more on data analysis and the combination of features. Our online submission results are based on fully connected layer with social information result. However, we find that profile with social features is much more effective than the method. And there are too few tag features in the method, and we don't use the tweet data and check-in data. In the future, we are going to use tag co-occurrence relationship to train a LINE vector to rich tag features in this subtask and split the users into two groups, the users that have tweets is divided into one group, the rest users is divided into another group, then construct a dual channel model to predict user tags.

For subtask two, we treat the task as a recommendation problem. We adopt a social-based collaborative filtering method. Since the number of users' common friends is small, the results of this method are not ideal. The collaborative filtering method can gain the similarity between users, so as to realize the friend recommendation. In this subtask, we only use social data. In the future, we can try to calculate similarity by using check-in data and profile data, and then combine multiple approach results to achieve the recommendation.

Acknowledgments. This research is supported by the National Key Research Development Program of China (No. 2016YFB1001103).

References

1. Farnadi, G., Tang, J., Cock, M.D., et al.: User profiling through deep multimodal fusion. In: Proceedings of the 11th ACM International Conference on Web Search and Data Mining, pp. 171–179. ACM (2018)
2. Farseev, A., Nie, L., Akbari, M., et al.: Harvesting multiple sources for user profile learning: a big data study. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pp. 235–242. ACM (2015)
3. Yan, C.L., Zhang, Y.Y.: Research of community discovery algorithm based on user tags. Sci. Technol. Eng. **11**(6), 1237–1240 (2011)
4. Harvey, M., Baillie, M., Ruthven, I., Carman, Mark J.: Tripartite hidden topic models for personalised tag suggestion. In: Gurrin, C., et al. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 432–443. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12275-0_38
5. Zhang, Y., Zhang, B., Gao, K., et al.: Combining content and relation analysis for recommendation in social tagging systems. Phys. A Stat. Mech. Appl. **391**(22), 5759–5768 (2012)
6. Sigurbjörnsson, B., Zwol, V.R.: Flickr tag recommendation based on collective knowledge. In: Proceedings of the 17th International Conference on World Wide Web, pp. 327–336 (2008)
7. Liu, Z., Shi, C., Sun, M.: FolkDiffusion: a graph-based tag suggestion method for folksonomies. In: Cheng, P.-J., Kan, M.-Y., Lam, W., Nakov, P. (eds.) AIRS 2010. LNCS, vol. 6458, pp. 231–240. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17187-1_22
8. Hu, J., Wang, B., Liu, Y., et al.: Personalized tag recommendation using social influence. J. Comput. Sci. Technol. **27**(3), 527–540 (2012)
9. Yamaguchi, Y., Amagasa, T., Kitagawa, H.: Tag-based user topic discovery using twitter lists. In: Proceedings of International Conference on Advances in Social Networks Analysis and Mining, pp. 13–20. IEEE (2011)
10. Wu, W., Zhang, B., Ostendorf, M.: Automatic generation of personalized annotation tags for twitter users. In: Proceedings of Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, pp. 689–692 (2010)
11. Mihalcea, R., Tarau, P.: TextRank: bringing order into texts. In: Proceedings of EMNLP, pp. 404–411 (2004)
12. Lappas, T., Punera, K., Sarlos, T.: Mining tags using social endorsement networks. In: Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 195–204. ACM (2011)
13. Chen, Y., Lin, L., Sun, C., et al.: A tag recommendation method for microblog users. Intell. Comput. Appl. **1**(3), 21–26 (2011)
14. Rangel, F., Rosso, P.: On the impact of emotions on author profiling. Inf. Process. Manage. **52**(1), 73–92 (2015)
15. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, pp. 43–52 (1998)
16. Sarwar, B., Karypis, G., Konstan, J., et al.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web, pp. 285–295. ACM (2001)
17. Li, B., Holtzman, H.: Online friend recommendation through personality matching and collaborative filtering. In: Proceedings of the 5th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, pp. 230–235 (2011)

18. Agarwal, V., Bharadwaj, K.K.: A collaborative filtering framework for friends recommendation in social networks based on interaction intensity and adaptive user similarity. Soc. Netw. Anal. Min. **3**(3), 359–379 (2013)
19. Lin, K., Chen, Y., Li, X., et al.: Friend recommendation algorithm based on location-based social networks. In: Proceedings of IEEE International Conference on Software Engineering and Service Science, pp. 233–236. IEEE (2017)
20. Tang, J., Qu, M., Wang, M., et al.: LINE: large-scale information network embedding. In: Proceedings of the 24th International World Wide Web Conference, pp. 1067–1077 (2015)