Activity recommendations based on heterogeneous social networks

Xiaoying Qi, Wen Yu
Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China
18232386357@163.com
qixiaoying@bupt.edu.cn

Abstract. EBSN (event-based Social Network) is a Heterogeneous Network, namely, H-EBSN (Heterogeneous event-based Social Network), which contains both online interactive information and offline interactive information. How to effectively use the interaction between online and offline to recommend activities is one of the hot issues at present. The following work has been done for the activity recommendation of H-EBSN:
(1) Construct H-EBSN, extract all link-roads and obtain user interest through link-roads’ characteristics.
(2) Analyze H-EBSN meta path, take the path number of extracted link-roads as feature vector, establish user interest model, and use generalized regression neural network to solve the model.
(3) The training set was constructed by using the movie-watching activity data set, and the generalized regression neural network was used to solve the model.
Experiments show that the activity recommendation method based on H-EBSN proposed in this paper has a good recommendation effect.

1. Introduction
Social networks can be divided into two main research scenarios: homogeneous networks[1] and heterogeneous networks[2]. A homogeneous network is a relatively single network line consisting of only one or two entities and one relationship. For example, if we only study the attention relationship between users, there will be only one entity "user" and one relationship "attention relationship". However, with the increasing complexity of social networks, considering only a single entity and relationship can not reflect the current situation well[3-4]. Moreover, the heterogeneity and complexity of social networks make it closer to the social networks in real life, which will also help to excavate more information and generate greater value for research. However, because heterogeneous networks contain different types of information and relationships, it becomes difficult to excavate truly valuable content[5]. As mobile Internet technology has become popular in recent years, a new type of heterogeneous social network, H-EBSN (Heterogeneous event-based Social Network) appears, such as Meetup, Plancast and EventBrite. In this kind of network, event organizers can publish spatial information of social events (such as event convening place and holding time) and event attributes on the social media platform, and then the social media platform will push different social events to potential participants. In short, event-based social networking platforms mainly link online users with offline events, and generate corresponding social services such as activity planning, activity recommendation and activity prediction, while pushing online events to potential users[6-8]. The "activity" in this article is the "event" in the social network.
The recommendation based on H-EBSN in this paper not only considers the attention relationship, but also integrates the content characteristics of users. In addition, the study on user behavior in H-EBSN is also conducive to further excavating activities that users are interested in, and improving recommendation results by excavating and integrating multi-channel features.

2. EBSN heterogeneous network construction

2.1. Basic characteristics of H-EBSN

The H-EBSN studied in this paper is a special Heterogeneous Social Network, which is composed of a variety of different types of nodes and different relationships among them. The structural change of the network depends largely on social media and social applications which means that people construct social networks based on current technologies and requirements, and extract the required heterogeneous node information[9]. For example, the H-EBSN of movie-watching activities includes three different types of nodes---user activity, activity tag and social relations (concern or be concerned) between user nodes, participation relations between user nodes and active nodes, and labeling relations between active nodes and tag nodes.

2.2. Construction of H-EBSN

In this paper, activity recommendation is carried out based on H-EBSN, and the H-EBSN which composed of three nodes, user, activity and activity tag, should be constructed first[10]. Category node VC, user node VU, resource node VR and tag VT, as well as the link relationship between each node TR, RT, TU, UT, UU, UR, RU, TC, CT, CU, UC, CR, RC constitute a heterogeneous network. In this article, $\lambda$ is defined as the shifting weight between different types of nodes (see Figure 1). Specifically, $\lambda_{UU}$ stands for the weight that a user chooses (concern) a user, or how much a user concerns about other users. $\lambda_{RT}$ stands for the weight of a resource selected (belonging) tag, or the width of the resource node in the tag node. The transfer weight $\lambda$ can also be seen as the degree of importance between the two types of objects.

Figure 1 Heterogeneous network model

H-EBSN includes three types of nodes: User, Activity and Tag: $V_U$ represents User node, $V_A$ represents Activity node, and $V_T$ represents Tag node. The H-EBSN model is shown in Figure 2.
2.3. Meta-path selection of H-EBSN

After the topology of the network is obtained, the meta-path is selected and then the meta-path with practical significance is screened out. This part mainly consists of three parts: network topology analysis, meta-path traversal extraction and meta-path screening. Network topology analysis mainly includes:

(1) Analyze and obtain node types in H-EBSN network topology.
(2) Analyze and obtain the node relationship in the network topology.

The core of meta path traversal extract part is to build a path extraction algorithm. The selection of meta-path is to filter out the non-meaningful meta-path from the extracted meta-path, because the analysis of the non-meaningful meta-path not only wastes a lot of calculation time, but also the calculated results can not be used to draw conclusions. Therefore, it is necessary to select meaningful meta-paths to analyze users’ interests and preferences. For example, in the movie-watching data set, \( U \rightarrow A' \rightarrow T \rightarrow A \) is a meaningful meta-path, which means that the activity A under the tag T as same as activity A’ is of interest to the user, so this meta-path is a link-road with practical significance. In the data set, there is no linked data of \( U \rightarrow T \rightarrow A \), so this meta-path has no practical significance.

The extraction algorithm of meta-path is as follows: first, according to the given H-EBSN topology, specify the shortest length \( l_{\text{min}} \) and longest length \( l_{\text{max}} \) of the meta-path, and then traverse the meta-path from the initial node F to the terminating node L. In the traversal process, the node N connected to the initial node F is first traversed, and then the nodes linked to node F, node L, and node N are traversed by f-N-L for the second time, and the nodes meeting the requirements of \( l_{\text{min}} < \text{node path length} < l_{\text{max}} \) are recorded and output, and finally filter out the non-meaningful meta-path.

The meta-path selection in H-EBSN is taken as the basis for recommendation, and the link-road relationship between heterogeneous nodes is collectively referred to as the meta-path[11]:

(1) \( U \rightarrow U' \rightarrow A \) means that there is a concerned user node U’ between user node U and activity node A, that means, the friend U’ that user U concerns participates in the activity A. So the user U is likely to participate in the activity A.

(2) \( U \rightarrow A' \rightarrow T' \rightarrow A \) means that there is another activity node and a label node in the path between the user node and the activity node, that means, user U watched the activity A’ with the same label as activity A, so user U is more likely to be interested in activity A.

It can be seen from the H-EBSN structure that user U is the initial node, activity A is the recommended termination node, nodes connected to user U are U and A, nodes connected to activity Node A are U and T, and nodes connected to tag node T are A.

After all meta-paths are extracted, the meaningful meta-paths are screened out and the non-meaningful meta-paths are filtered out.
As the length of the meta-path increases, the user node becomes less connected to the recommendation node and less interested in the recommendation node. The shorter the length of the meta-path, the closer the relationship between the user and the recommended node, and the greater the user's interest in the node. The output of user node to activity node can be used to represent the frequency of user participation, and the output of active node to user node can be used to represent the frequency of activity participation.

3. EBSN heterogeneous network recommendation

3.1. Framework of recommendation

This paper will study the method to obtain the meta path in H-EBSN, calculate the path feature weight as the feature vector, and establish the user interest model. At the same time, the topology structure of H-EBSN[12-13] needs to be analyzed in order to screen the meta-path with practical significance. Therefore, this paper extracts the network path feature vector based on the network topology structure diagram given by experimental data, and builds the recommendation model and solution through the extracted meta-path and analysis data.

In the part of network topology construction, this paper constructs the network according to the required nodes and user data obtained. Different from collaborative filtering and content-based recommendation algorithms[14-15] and complex network link prediction methods, the main work of this paper is summarized as follows:

1) Use traversal to extract the H-EBSN link-road, mining the meaningful meta-path, calculate the meta-path weight according to the degree of output between S node/nodes of each type, and substitute the calculated feature vector into the model for analysis. Unlike link-road prediction, which only considers the local path and the whole path, this paper selects the meta path with practical significance and uses the computed eigenvector to explore the user interest.

2) Calculate the path number of each extracted link-road, divide the product of the total number of heterogeneous nodes by the total number of paths as the feature vector, establish the user interest model, and solve the input feature value of each meta-path through GRNN (General Regression Neural Network) to get the final recommendation model.

3.2. Structure of GRNN

In this part, we establish and solve the recommendation model. After the data processing, the generalized regression neural network is used to train the training set data, establish the recommendation model, and finally verify the test set data to evaluate the recommendation results.

GRNN is a three-layer forward network, the input layer, the hidden layer and the output layer, including a radial base layer and a linear layer, as shown in Figure 3. The input layer does not participate in the actual operation, and only feeds the sample variables into the hidden layer. The number of neurons in the hidden layer is equal to the number of training set samples. The Euclidean distance function (represented by ||dist||) is used in this layer to act on the weight function to calculate the distance between the network input and the weight LW_{1,1} of first layer. b1 is the threshold value of the hidden layer. The transfer function of the hidden layer is the radial basis function, and the Gaussian function is usually used as the transfer function of the network. The third layer of network is linear output layer, and it’s weight function is normalized dot product weight function (expressed in nprod). It computes network vector as n2, and it’s elements are obtained by dividing the dot product of each row element by the sum of each element of vector a2. The result n2 is provided to the linear transfer function a2 =purelin(n2) to compute the network output[16].
Figure 3 Structure of GRNN

In the training of GRNN, the number of radial basis neurons and linear neurons is the same as the number of input vectors in the input training sample. The purpose of network training is to generate appropriate weight matrix LW1,1 and LW2,1 and threshold vector b1.

3.3. Use GRNN to build recommendation model
In this paper, the weight of the relationship between the initial node and the termination node is calculated through each meta-path as the xn of the input layer (As shown in Figure 4), the relationship between the actual initial node and the termination node is taken as the output layer. The GRNN model is used for training, and the recommendation effect is evaluated through the test set.

Figure 4 Each layer of GRNN

4. Experiments

4.1. Data preprocessing
Experiments use the data set of movie-watching activities. There are three kinds of node types and each type contains thousands of nodes. The relationship between the nodes can reach more than thousands, so to build a H-EBSN need to spend a lot of time, looking for meta path between user nodes and activity nodes is complex too. For example, the path of U - U' - A (assuming the friends concerned participates in the activities), due to the complex relationship between nodes, the user node U may pay attention to a lot of users U', and for each user U', he or she may be involved in many activities A too. So for this simple meta-path, the amount of computation required is very large.

The purpose of the experiment is to use the data set to construct the H-EBSN structure, then build a model based on the H-EBSN structure and make recommendations to verify the effectiveness of the proposed method. In order to reduce the complexity of the network structure, it is necessary to preprocess the data set and remove the redundant edges between nodes. The data processing results are as follows: delete 1063 independent user nodes (those who do not follow friends or participate in
activities) and 2347 independent activity nodes (those who do not participate or are not tagged by users), and finally form the user activity data consisting of 8039 users, 5542 activities and 1121 tags.

4.2. Solution of recommendation model for data set of movie-watching activity

The experimental data in this paper are derived from the recommended data set for movie-watching activities. The data form obtained after MATLAB data preprocessing is shown in Table 1, which is divided into training set data and test set data. In this data form, 0 means that the meta-path is not connected, that is, $U \rightarrow A$ value of 0 means that the user does not participate in the activity. $U \rightarrow U' \rightarrow A$ is 6, indicating that 6 of the friends followed by the user have participated in the activity, and the meta-path $U \leftarrow U' \rightarrow A$ is 3, indicating that 3 of the users following the user have participated in the activity.

| Meta path $U \rightarrow U' \rightarrow A$ | Meta path $U \leftarrow U' \rightarrow A$ | Practical situation $U \rightarrow A$ |
|--------------------------------------|--------------------------------------|----------------------------------|
| 0                                    | 1                                    | ......                           |
| 2                                    | 0                                    | ......                           |
| ......                                | ......                                | ......                           |
| 0                                    | 0                                    | ......                           |
| 6                                    | 3                                    | ......                           |

Multiplying each meta-path by weight, then input them to each input layer of GRNN. The value of path $U \rightarrow A$ was used as the output layer, that means, the actual movie watched by the user. Figure 5 shows the comparison between the user's participation in 100 activities and the prediction of GRNN(True value: 1 means yes, 0 means no;Predicted value: 1 is recommended, 0 is not recommended).

Figure 5 Comparison of test set predicted results

In this paper, four meta-paths, $U \rightarrow U' \rightarrow A$; $U \leftarrow U' \rightarrow A$; $U \rightarrow A' \rightarrow U' \rightarrow A$ and $U \rightarrow A' \rightarrow T' \rightarrow A$, are combined as each input layer in GRNN, and the accuracy of training results is shown in figure 6. We find that the accuracy calculated by model 4, 7 and 9 is better than the traditional collaborative recommendation algorithm, while Model 4 is a combination of meta-path 1, 2, 3 and 4 and model 7 is a combination of meta-path 2, 3 and 4, and Model 9 is a combination of meta-path 3 and 4. Therefore, the recommendation accuracy based on H-EBSN social network is relatively high under multiple meta-paths, and the recommendation results are relatively good.
5. Conclusions

According to the algorithm of meta-path extraction, the meta path of H-EBSN social network is screened with practical significance, and the filtered meta-path is weighted. Finally, the experimental data is substituted into GRNN to evaluate the recommended effect.

The main works of this paper are as follows:

(1) Proposed a method of extracting meta path with practical significance. For the network topology given by users, we analyze the relationships between network nodes, and extract all the relationship links in the network by using meta-path extraction algorithm, and then screen out the relationship links with practical significance.

(2) After analyzing the meta-path in H-EBSN, the meta-path with practical value is selected as the recommendation basis. The path number of each extracted relational link is calculated, and the result is input as the feature vector to build the user interest recommendation model, then the generalized regression neural network is used to solve the model.

References

[1] Deng Cai, Zheng Shao, Xiaofei He, et al. Mining hidden community in heterogeneous social networks[P]. Link discovery, 2005, :58-65.

[2] Chaobo He, Yong Tang, Guohua Chen, Mei Liu, Linlin Wu. Potential friend recommendations for large-scale social networks[J]. Journal of Hefei University of Technology (Science), 2013,04: 420-424. (In Chinese)

[3] Jie Tang, Wenguang Chen. In-depth analysis and mining of big social data[J]. Science Bulletin, 2015, Z1:509-519. (In Chinese)

[4] Aatook S, Brown S A, Rolf J. Forming an intention to act on recommendations given via online social networks [J]. European Journal of Information Systems, 2015, 24(1): 76-92.

[5] Jing Zhang, Jie Tang, Bangyong Liang, et al. Recommendation over a Heterogeneous Social Network [J]. IEEE Conference Publications, 2008, 71: 309-316.

[6] Xuecheng Yang, Xiaobo Tao, Xin Yue. Simulation modeling and Empirical Analysis of double finite Heterogeneous social networks[J]. Journal of Beijing University of Posts and Telecommunications, 2015, S1:121-124.(In Chinese)

[7] Xiyue Gao, Jun Chen, Zexing Zhan, et al. Learning Heterogeneous Information Network Embeddings via Relational Triplet Network[J]. Neurocomputing, 2020, 412:31-41.
[8] Hailu Yang, Jianpei Zhang, Jing Yang, Heterogeneous network communities were discovered by 2-hop random walk [J]. Journal of Harbin Engineering University, 2015, 12: 1626-1631. (In Chinese)

[9] Yen-Kai Wang, Wei-Ming Chen, Cheng-Te Li, et al. Identifying smallest unique subgraphs in a heterogeneous social network [C] // IEEE International Conference on Big Data. IEEE, 2015.

[10] Mukul Gupta, Pradeep Kumar, Bharat Bhasker. DPRel: A Meta-Path Based Relevance Measure for Mining Heterogeneous Networks [J]. 2019, 21(5): 979-995.

[11] SUN Y, HAN J, YAN X, et al. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks [J]. Proceedings of the VLDB Endowment, 2011, 4(11): 992-1003.

[12] Mukul Gupta, Pradeep Kumar. Recommendation generation using personalized weight of metapaths in heterogeneous information networks [J]. European Journal of Operational Research, 2020, 284(2): 660-674.

[13] Hongzhi Liu, Zhengshen Jiang, Yang Song, et al. User preference modeling based on meta paths and diversity regularization in heterogeneous information networks [J]. Knowledge-Based Systems, 2019, 181.

[14] Ke Chen, Quan Zou, Zhiping Peng, Wende Ke. Collaborative sorting of Friend recommendation algorithms in heterogeneous social networks [J]. Minicomputer system, 2014, 06: 1270-1274. (In Chinese)

[15] Kehan Chen, Panpan Han, Jian Wu, A heterogeneous social network recommendation algorithm based on user clustering [J]. Journal of Computer Science, 2013, 36(02): 349-359. (In Chinese)

[16] Kaibo Zhou, Dengzhi Yu, Zhikai Lin, et al. Anode effect prediction of aluminum electrolysis using GRNN [C] // Chinese Automation Congress. IEEE, 2016.