Future Generation Computer Systems 78 (2018) 462–473

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

FLOWER: Fusing global and local associations towards personalized social recommendation

Mingxin Gan [a,*], Rui Jiang [b]

[a] Department of Management Science and Engineering, Donlinks School of Economics and Management, University of Science and Technology Beijing, China
[b] Department of Automation, Tsinghua University, Beijing, China

HIGHLIGHTS

- We found potential social relations are related with common connections.
- We demonstrated the necessity of fusing global and local associations.
- We developed a novel method FLOWER to integrate global and local associations.
- Methods based on FLOWER significantly outperform others in social recommendation.
- We demonstrated the effectiveness of FLOWER on five social networks.

ARTICLE INFO

Article history:
Received 25 June 2016
Received in revised form
19 January 2017
Accepted 15 February 2017
Available online 20 February 2017

Keywords:
Social recommendation
Fisher's combined probability
Network motif
Global association
Accuracy
Facebook

ABSTRACT

In big data era people are dependent on a variety of social media to manage their social circles. Many online social networks employ social recommendation as an increasingly important component. Although global and local recommendation methods have achieved remarkable success, current studies seldom consider to play advantages of both in social networks. To demonstrate the effectiveness of incorporating local methods to global ones, we first investigated associations between triangular motifs and existing social relations and found that potential links are related with common relations. Further, we analyzed correlations of all methods and clustered them and found obvious strong correlations among methods with same type, which demonstrated the necessity of taking advantages of both. Consequently, we proposed a novel method FLOWER which resorts to Fisher's combined probability test to systematically calibrate statistical significance of global and local associations. FLOWER utilizes information of social relations in both local and global scopes, which are less correlated with each other, and therefore imply possibilities of different aspects for a candidate link. We demonstrated the effectiveness of FLOWER by considering each possible pairwise combination of six global approaches with two local methods and performing 10-fold cross-validation experiments on five real social network datasets (Facebook, Last.fm, Epinions, HEP-PH and Delicious). Results show that FLOWER-based methods significantly outperform either their global or local components in accuracy and retrieval performance.

1. Introduction

The last few years had witnessed an explosion of information caused by the exponential growth of the Internet [1] and World Wide Web [2], which confronted us with information overload and brought about an epoch of big data. In such an era, recommendation systems, as an important way to improve the efficiency of information acquisition, have been greatly alleviating information overload by providing personalized nomination of resources [3–6], such as movies [7], bookmarks [8], music [9,10], videos [11,12], news [13,14], online advertising [15], microblogs [16], electronic products [17] and locations [18,19] to enhance online sales [20], increase sale diversity [21] and benefit e-commerce merchants [22].

In such an era, people have been generating overwhelming resources in a variety of social media and communities, and relying on such information to manage their personal social circles [23]. Consequently, such online social networks as Facebook, Twitter, Last.fm, Epinions, and YouTube, not only provide novel ways for people to communicate and share their resources and opinions

* Corresponding author.
E-mail addresses: ganmx@ustb.edu.cn (M. Gan), ruijiang@tsinghua.edu.cn (R. Jiang).

http://dx.doi.org/10.1016/j.future.2017.02.027
0167-739X/

with each other, but also serve as a platform to establish new relationships among people. Furthermore, to promote the evolution of social networks and meet the demand of interactions and communications, such online social networks as Google+, Facebook and Twitter, have provided the functionality of suggesting friends to people through utilizing their personalized information [24,25], making friend recommendation an increasingly important component in studies of social networks [26].

In general, there are two branches of study in recommender systems [27,28]: ranking [29] and rating prediction [30,31]. Social recommendation methods suggest possible connections to a target person and motivate the person to establish certain relationships. Therefore, social recommendations try to solve the ranking problem, known as Top-N recommendations in social networks, which takes a person and a set of candidate friends as input and produces a ranking list of the candidates as output [32]. Up to now, many efforts have been devoted to social recommendations. Specifically, representative methods for Top-N social recommendation are divided into three categories.

The first category includes methods considering existing relations simply to calculate ranking scores of candidates. Since existing relations play the most important role, there are many success collaborative filtering approaches [27]. Classical collaborative filtering resorts to historical relationships to infer future probabilities that people establish relations [33,34], such as ItemKNN [35,36], UserKNN [27] and association rules-based methods [37]. Global-paths based methods as random walk and network propagation consider the propagation of people's preference or connections [38–42]. Typical methods include network-based methods [28,43–45] and random walk with restart [46,47].

Nevertheless, existing relations are not the only resource that characterize people's social preference. Therefore, the second category includes methods that explore features hidden behind visible relations to investigate potential possibilities of social relations. Methods based on matrix factorization make recommendations relying on underlying factors in social networks [34,48,49]. The basic idea of matrix factorization methods is to extract features in a somewhat low dimension, and then use these features to calculate predictive scores for possible associations. Typical models include non-negative matrix factorization (NMF) [50,51], SVD [52]. Since the above two categories of methods explore connections throughout the entire network, we denote them as global methods.

As the third category, methods based on local neighbors consider network motifs [53] as important components in social recommendation. Such classical methods as common neighbors and the Jaccard index make recommendations by assigning a score reflecting numbers of triangular-motifs between two persons [53–55]. The '*triadic closure*' rule indicates that, if two persons share a common friend, there is an increased probability of a relation between them in social networks in future [53]. Therefore, we denote such methods as local methods.

Although both global and local approaches have achieved their success in social recommendations, adopting local correlations alone may disregard the propagation of connections [56] while utilizing global associations singly may disregard the essential '*triadic closure*' rule. Based on the understandings, the objectives of the study are to answer the following questions:

- What is the correlation between triangular motifs and social relations?
- Is it necessary to take advantages of both global and local methods for social recommendations?
- What roles global and local associations play in social recommendations?
- How to design a hybrid method to effectively fuse both global and local associations to enhance social recommendation performance?

Our contributions in this study are summarized as below. First, we analyzed correlations between triangular motifs and social relations, which suggests that if a person has a large number of social relations, the connections among his friends will be large as well. Second, we demonstrated the necessity of taking advantages of both global and local methods by calculating Pearson correlation coefficient for each pair of methods and cluster all methods using hierarchical clustering. Results show obvious strong correlations among methods with same type. Third, based on the above investigations, to achieve a balance between global and local associations, we propose a novel method FLOWER (**F**using G**L**obal and L**O**cal Associations to**W**ards P**E**rsonalized Social **R**ecommendation) to fuse global and local associations. Specially, a rigorous statistical model is used to calibrate the statistical significance of these associations, which resorts to Fisher's combined probability test to obtain a single p-value as the final score.

We demonstrate the effectiveness of FLOWER by considering each possible pairwise combination of six global with two local recommendation methods and performing 10-fold cross-validation experiments on five real social network datasets (Facebook, Epinions, HEP-PH, Last.fm and Delicious). Results show that approaches based on FLOWER, remarkably outperforms either their global or local components in both accuracy and retrieval performance, indicating its promising future for social recommendation.

2. Related work

Representative methods for Top-N social recommendation are divided into three categories. The first category includes methods considering existing relations simply to calculate ranking scores of candidates. The second category includes methods that explore features hidden behind visible relations to investigate potential possibilities of social relations. The above two categories of methods explore connections throughout the entire network, we denote them as global methods. The third category includes methods based on local neighbors considering network motifs as important components in social recommendation and therefore, we denote such methods as local methods.

2.1. Global methods

Global methods consider the propagation of people's preference or connections [38,40]. Such approaches include random walk with restart (RWalk) [46], diffusion kernel (DKernel) [57–61], probability spreading (ProbS) [62], network propagation [38,39,42], classical collaborative filtering [33,63,64] and matrix factorization [34,48–50,52]. We categorized current state-of-the-art methods into 3 types: methods considering historical ratings, methods incorporating hidden feature layers, and methods incorporating features.

Methods considering historical ratings

Methods considering only ratings resort to using known social relations to calculate ranking scores of candidates. The most commonly employed method, collaborative filtering based on user similarity (USim) [27], calculates a user similarity matrix from historical ratings based on the cosine measure.

Classical collaborative filtering methods resort to historical social relations to infer future probabilities that people will establish connections [33,63], i.e. user-similarity-based collaborative filtering (USim) [27], which weights the preferences of persons according to their similarities with a target person and mixes the preferences to obtain discriminant scores for candidate friends [27]. Then, taking similarities as weights of connections, USim searches for all simple connections between people and calculates the summation of the weights of these connections to characterize the association between them. Given the similarity matrix $\mathbf{S} = (s_{uf})_{n\times n}$,