A Framework for APP Discovery Based on Relation Mining

Xiao-li LIU, Long-xin LIN* and Wei-zhen JIANG
College of Information Science and Technology, Jinan University, Guangzhou, China
*Corresponding author

Keywords. APP discovery, Relation mining.

Abstract. In this paper, we propose a framework of APP discovery for a given user based on APP relation mining. First, Personalized Trust Community (PTC) is constructed based on complex network techniques with “APP-User” data. Then we discover fresh APPs based on APPs latent relations which is mined through PTC analysis. We plan to experiment our framework with the data crawled from APP store.

Introduction

Appropriate APPs can help mobile users improve work efficiency. Nowadays, searching is the top app discovery tool. The study show that 63 percent of Android and iOS users have utilized search to discover fresh apps in their respective app stores [1]. Nonetheless, when we want to search a fresh app, we will find that “There Is An App For That … Oh Wait, There Are 100 Apps For That”. So many similar apps in APP store so that we don’t know which one to choose. Forbes AllBusiness published an article which present a good explanation the importance of APP discovery [2].

There are many factors which can influence the searching result, such as APP Miscategorization, wrong APP description and so on. It points that the Google game APP miscategorization can be as high as 3.32-11.08 percent [3]. Even with the correct categorization, the APP behavior is not consistent with APP description now and then. Hence, some researchers dedicate to check APP behavior against APP description [4] [5]. Moreover, it is hard to get their personalized information when they search for fresh APPs.

At present common method for APP discovery is recommendation from third parties or social relations. Such recommendation method have better effort for APP discovery than searching. In [6] the context-aware method is often used for recommendation. In [7], inter-App relations (i.e., intent-based, semantics correlation based, and similarity-based ones) are launched for recommendation. These works have achieved good results. However, privacy protection is neglected. In [8], social network is collected using built-in sensors. The explicit feature from sensing readings of built-in sensors and the implicit feature from App usage relations in [9]. Even though, some recommendation model utilizes the diverse range of implicit mobile data available in a fast and scalable manner [6].

To sum up, we need further method for APP discovery by combining the latent relations without violating user privacy. In this paper, we will build a personalized trust community model for app discovery which may imply latent inter-app relations based on users’ habits with no privacy exposure.

Related Work

In this section, we summarize the research status about relation extraction about APP discovery and user preference mining method. And also, current privacy reserved approach for APP discovery is presented.
User Social Relation

Social relation is a kind of resources and it denotes the relationship between persons. In most cases, two persons may have some common features if they are social related. Composite social network is used for predicting mobile APP installation [8]. In [11], social users’ contextual information is used for service quality evolution. In [12] a social network based service recommendation method is presented. These research results show that social network play a big role for APP prediction and have good effort.

As for research method about social relations extraction, too many techniques are adopted, such as matrix factorization, entropy theory, machine learning and so on. In [12] a matrix factorization method is used to evaluate the trust degree of social-related users. In [11], entropy theory is used for calculating user’ confidence.

Inter-APP Relation

Inter-APP relations research mainly divides into two classes: functionality concerned relation and personalized relation. Functionality concerned relation refers to the execution sequence, eg., one app needs another app to cooperate to fulfill its work. It has no relevance with user’s habits or features. Whereas, personalized APP relation means usage sequence of APPs for given user. It usually involves user’s preference information. As definition in [7], functionality concerned relations are represented by Global App Network (GAN) which is constructed from APP store and personalized APP relations are represented by Personal App Network (PAN) which is constructed from APP usage log. Personalized relations are usually studied by analyzing user’s log. In [9] implicit transition relation among Apps is elicited by APP usage session analysis. In [13] semantic relations between apps are defined using ontology, and also the similar social members is identified.

Functionality concerned relations are usually studied by analyzing app’s contexts or reviews [14]. In [15] relation about functionality and topic between APPs are represented by cooperation relation. APP relationship is calculated via review analysis with iterative process, performs better than some baseline method by a large margin [14]. APP network is studied through cluster analysis with Apple’s App Store [16]. The research result aims to provide implications for mobile ecosystem participants. APP related features are studied for recommendation [17], in which user’s history, app metadata, and the recommendation scores are involved for analyzing.

Above all, common researching methods about inter-APP relationship extraction include text mining [16], probabilistic matrix factorization [17], and so on.

PTC Definition (Preliminary)

The APP usage habit is influenced by the user’s career, hobby, age, and so on. Let’s makes this a little bit clearer with an example: A young teacher which like surfing has different APP usage with an old teacher who like climbing. Although the young teacher and the old teacher may have social relations, they may use most different APPs. What we want to do is to construct personalized trust community for a given person. In [18, 19] the trustworthy community is constructed for service recommendation, which can be used as reference of our work.

Latent Inter-App Relations

At this point, the APP relations are not just about the function interdependency, but the co-occurrence of different APPs. If two APPs appear in different users’ mobile phone simultaneously, we affirm that the two APPs are related to some extent. In practice, if two users have high similar characteristics, they usually use more than one similar APPs. Usually, there will be several APPs which are related under our definition. Therefore, we need to define the number of APPs and strength of their relations in a closely related group.
In reality, there is no clear bound among different APP groups. What we should do is to find out the substantial relations and get rid of the interference of non-personalized APPs. For instance, the popular APPs and queer APPs should be eliminated due to their preference-irrelevance. Usually, the inter-related APPs may have function interdependency to some degree.

**Personalized Trust Community**

We define Personalized Trust Community (PTC) means similar user group which have similar latent inter-APP relations to a given user. Among PTC, the users and the latent similarity relationship are stored. What we want to do is to find almost identical users for PTC construction for a given user. If someone use the same APP with a given user, he will be added in the raw PTC of given user.

PTC can be best described by weighted networks with a diversity of interactions between nodes measured by the weights of the edges. In PTC network, there may be various user nodes which have different weights. The nodes weights denote the strength of the relation between the user node and given user. The weights of edge denote their strength of the relation between the two nodes.

**Framework**

Our aim is to construct the PTC and discover APPs based on PTC. In this section, we first give the workflow of the framework.

**Overview**

Figure 1 show the implementation framework of our method.

![Figure 1. Implementation framework for APP discovery.](image)

The process contains following three steps:

**Step 1**: APP-related user capturing. We will get the related users which use one or several common APPs with given user by review analysis. And also, the relations among users can be obtained according to the related users’ review. One step-depth mining is adopted for relation construction. The idea is that the relation will be constructed or strengthened if two users use one common APP.

**Step 2**: Relation refinement and evolution. The incidental related users are useless. That is to say, the weakly related users will be eliminated according to the weight of node and relations. The rules of evolution should be developed.

**Step 3**: Algorithm design for APP discovery. The APP discovery algorithm will be constructed based on PTC.

**PTC Construction**

PTC is constructed based on improved complex network. We use complex network analysis techniques for data representation, relation calculating and PTC construction.
Data Gathering

We use Two-layers-exploration method for related user acquisition from “user-APP” data. For a given user, we collect related information as follows:

- Used APPs. APPs are using by the given user.
- APP related users. The users who are using any one of the APPs in above step.
- Related users’ APPs. They mainly refer to the APPs which are used by related users.

Representation

A network is a mathematical representation of a real-world system and is defined by a collection of nodes (vertices) and links (edges) between pairs of nodes. We adopt weighted bipartite network for our data processing. We intend to build the network with the following rules:

- The users and APPs are represented by nodes with different marks.
- If an APP is used by a user, they will be connected with an edge.
- If two users use the same APP, they will be connected with an edge. The edge weight is equal to the numbers of the same APPs.

Figure 2 show the data representation for PTC formation. The above red nodes represent the users and the yellow ones represent APPs. The dashed edges show the usage similarity between users and solid edges show the usage relationship. Note that we do not record complete relationship among users and APPs. Only users centered information are represented by the weighted network.

Calculating for PTC Construction

Improved weighted bipartite network is adopted to represent PTC which contains weighted nodes and edges. The weighted nodes represent the users and the similarity degree to given user. The weighted edges represent the similarity between two linked users. The complex network techniques will be adopted for PTC construction and evolution, such as clustering, outlier detection [20] [21] and so on. The similar users use similar APPs. It is expected that the users in PTC have high similarity with the given user.

APP Discovery Based on PTC

We deem that these APPs in the same group have latent relations. What we want to do is to explore the latent relations. We can resort to complex network techniques for relations mining. Then, the fresh APPs can be obtained according to APP relations matching.

Implementation Setup

The goal in this experiment was to assess the APP discovery accuracy with our network based PTC construction framework using inter-APP relations.
Data Collection

It is difficult to gather all related users and APPs by investigation since unknown similar users are missed. In the initial phase, we want to get the related information with the help of APP reviews in APP store. A person may review an AP only if he deeply concern about the APP. Thus, we can get the strong correlation between the user and APP. We intend to user-APP relates from reviews in APP store.

In the empirical study, WanDouJia is selected as the data source of mobile Apps. It is a top-3 Android App Store in China with more than 3.2 million Apps in more than 20 categories as of October 2017, and compared with other App stores, it offers more powerful APIs which facilitate App crawlers to easily collect information of massive Apps. In view of APP real-time feature, we merely collect recently six month data for analysis. We randomly select one user as the given one.

Due to the timeliness of APP, we plan to crawl the data in one year. All the APP related reviews will be collected for experimentation, although most reviews have no connection with the given user.

Experimentation Design

The experimentation data contains recent 12-month APPs and reviews. We plan two ways for experimentation:

1) Success rate prediction of APP in last month. Take the first 11-month data as input to predict the APPs in last month, in which the recall and precision are taken into consideration. To do this, we pick one user randomly from 11-month data and get APPs in last month as a comparative data.

2) Hidden specific information of given user for discovery. We will pick out a given user and one APP of this user randomly. Then the APP and APP–related data is cut down. We experiment whether the hidden APP can be discovered.

The experimentation should be carried out for different random users by above two methods to get the average precision and recall. Now we are processing data, including data crawling and cleaning. Because of the time constraint, the experimentation have not completed.

Conclusion

It is hard to get their personalized information when they search for fresh APPs. There is no doubt that the search result is unsatisfactory since the personalized information is lacking. Personalized information should be taken into consideration for APP discovery, such as user preference, social relations and so on. Therefore, isolated study about APP itself cannot gain better result for recommendation.

In this paper, we propose a framework for APP discovery for a given user based on APP relation mining. First, we construct personalized trust community (PTC) with complex network techniques from APP-user related data. Then the APP is discovered through APP set matching between given user and similarity user in PTC. We plan to experiment our framework with the data in APP store. Due to the timeliness of APP, we plan to crawl the data in one year. All the APP related reviews will be collected for experimentation, although most reviews have no connection with the given user. PTC can also be used for malicious APPs filtering, preference mining and so on.

Acknowledgement

This work is partially supported by the National Science Foundation of China (Grant No. 61501207), Guangdong Provincial Scientific and Technological Projects (Grant No. 2016A010101018 and 2016A010119171), Guangzhou Scientific and Technological Project (Grant No. 201601010314 and 2014Y2-00133), Natural Science Foundations of Guangdong (No. 2016A030313104), the
References

[2] Haselmayr, M. App Discovery: Why Can't Anyone Figure This Out Yet? 2013. Forbes AllBusiness.