Personalized POI Recommendation on Location-Based Social Networks

by

Huiji Gao

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved November 2014 by the
Graduate Supervisory Committee:

Huan Liu, Chair
James Caverlee
Guoliang Xue
Jieping Ye

ARIZONA STATE UNIVERSITY

December 2014

ABSTRACT

The rapid urban expansion has greatly extended the physical boundary of our living area, along with a large number of POIs (points of interest) being developed. A POI is a specific location (e.g., hotel, restaurant, theater, mall) that a user may find useful or interesting. When exploring the city and neighborhood, the increasing number of POIs could enrich people's daily life, providing them with more choices of life experience than before, while at the same time also brings the problem of "curse of choices", resulting in the difficulty for a user to make a satisfied decision on "where to go" in an efficient way. Personalized POI recommendation is a task proposed on purpose of helping users filter out uninteresting POIs and reduce time in decision making, which could also benefit virtual marketing.

Developing POI recommender systems requires observation of human mobility w.r.t. real-world POIs, which is infeasible with traditional mobile data. However, the recent development of location-based social networks (LBSNs) provides such observation. Typical location-based social networking sites allow users to "check in" at POIs with smartphones, leave tips and share that experience with their online friends. The increasing number of LBSN users has generated large amounts of LBSN data, providing an unprecedented opportunity to study human mobility for personalized POI recommendation in spatial, temporal, social, and content aspects.

Different from recommender systems in other categories, e.g., movie recommendation in NetFlix, friend recommendation in dating websites, item recommendation in online shopping sites, personalized POI recommendation on LBSNs has its unique challenges due to the stochastic property of human mobility and the mobile behavior indications provided by LBSN information layout. The strong correlations between geographical POI information and other LBSN information result in three major human mobile properties, i.e., geo-social correlations, geo-temporal patterns, and

geo-content indications, which are neither observed in other recommender systems, nor exploited in current POI recommendation. In this dissertation, we investigate these properties on LBSNs, and propose personalized POI recommendation models accordingly. The performance evaluated on real-world LBSN datasets validates the power of these properties in capturing user mobility, and demonstrates the ability of our models for personalized POI recommendation.

DEDICATION

I dedicate my dissertation work to my loving parents, Haodu Gao and Lixing Zhang, for making me be who I am!

I also dedicate this dissertation to my precious wife, Ye Wu, for supportng me all the way! Without her help and encouragement it simply never would have been.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Huan Liu, for his consistent support during my Ph.D. study. Dr. Liu helped to cultivate my research strength, provide invaluable advices in the last five years. Working with him, I learned to write my first paper and addressed many challenging problems. More importantly, I learned from him the work attitude and disciplines that I will benefit all my life. I could not have imagined having a better advisor for my Ph.D study. Dr. Liu is more of a mentor and life long friend, than a professor.

I would like to thank my thesis committee, Dr. James Caverlee, Dr. Guoliang Xue, and Dr. Jieping Ye, for the assistance, encouragement, and insightful comments that they have provided at all levels of my research.

I would like to thank Dr. Jalal U. Mahmud (IBM Research) and Dr. Kun Liu (LinkedIn), for providing me the summer internship opportunities in their groups and mentoring me on diverse exciting research projects.

I would like to thank my colleagues at the Data Mining and Machine Learning Lab for their helpful suggestions and support. It is a great pleasure working with them, particularly, Jiliang Tang, Xia Hu, Lei Tang, Zheng Zhao, Geoffery Barbier, Shamanth Kumar, Fred Morstatter, Ali Abbasi, Reza Zafarani, Pritam Gundecha, Isaac Jones, Zhuo Feng, Suhas Ranganath, and Salem Alelyani.

I would like to thank the Office of Naval Research (ONR) and Arizona State University for their continous support through the grant (No. N000140810477, No. N000141010091, and No. N000141110527) and Graduate Education Dissertation Fellowship during my entire Ph.D. study.

Last but not the least, I would like to thank my family for the support they provided me through my entire life and in particular, I must acknowledge my wife Ye Wu, without whose love and encouragement, I would not have finished this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Chapter 1

INTRODUCTION

The rapid growth of cities has developed an increasing number of points of interest (POIs), e.g., restaurants, theaters, stores, hotels, to enrich people's life and entertainment, providing us with more choices of life experience than before. People are willing to explore the city and neighborhood in their daily life and decide "where to go" according to their personal interest and the various choices of POIs. At the same time, making a satisfying decision efficiently among the large number of POI choices becomes a touch problem for a user. To facilitate a user's exploration and decision making, POI recommendation has been introduced by location-based services such as Yelp[1] and Foursquare[2]. However, such recommendation models are commonly based on majority users' preference on POIs, which ignore a user's personal preference. Comparing to visiting places that best fit a user's interest, visiting places against a user's taste may give him very terrible experience, especially in a situation when the user travels to a new place. Therefore, personalized POI recommendation is proposed to help users filter out uninteresting venues according to their own taste and save their time in decision making.

1.1 Background

Before the Web 2.0 era, analyzing user's mobility for personalized POI recommendation is infeasible even the mobile devices are widely adapted with large amount of cellphone-based GPS data available, as there is no indication of POI information from

[1] http://www.yelp.com
[2] http://foursquare.com

the GPS data other than longitude and latitude records. For example, we could observe a set of locations in terms of longitude and latitude pairs that a user has been to, while there is no easy way to figure out whether a specific pair of longitude and latitude is corresponding to a restaurant, or a hotel, or just a point on highway, since all these information are passively recorded by mobile devices.

With the developing of Web 2.0 technology, a number of location-based social networking services, e.g., Foursquare, Yelp, and Facebook Places[3], have emerged in recent years, making the study of personalized POI recommendation possible. Typical location-based social networking services maintain a large POI database and allow a user to "check-in" at a POI with his smartphone regarding to his current physical location. The user can also leave tips and share the "check-in" experience with his online friends, along with creating the opportunity to make new friends. According to a recent survey from the Pew Internet and American Life Project, over the past year 18% of smartphone owners use geosocial services to "check in" at certain locations and share them with their friends, while this percentage has risen from 12% in 2011 [102]. Such rapid growth has led to the availability of a large amount of user mobility data, promoting a new concept of online social media, namely location-based social networks (LBSNs).

Location-based social networks not only refer to the social connections among users, but also consist of the "location-based" context including geographical check-in POIs, check-in time stamps, and check-in related content (e.g., tips, comments, POI descriptions, etc.), as shown in Figure 1.1. Compared with other online social networks that consist of user activities interacting with the virtual world, LBSNs reflect a user's geographical action in the real world, residing where the online world and real world intersect, therefore bridging the gap between the real world and the vir-

[3]http://www.facebook.com/about/location

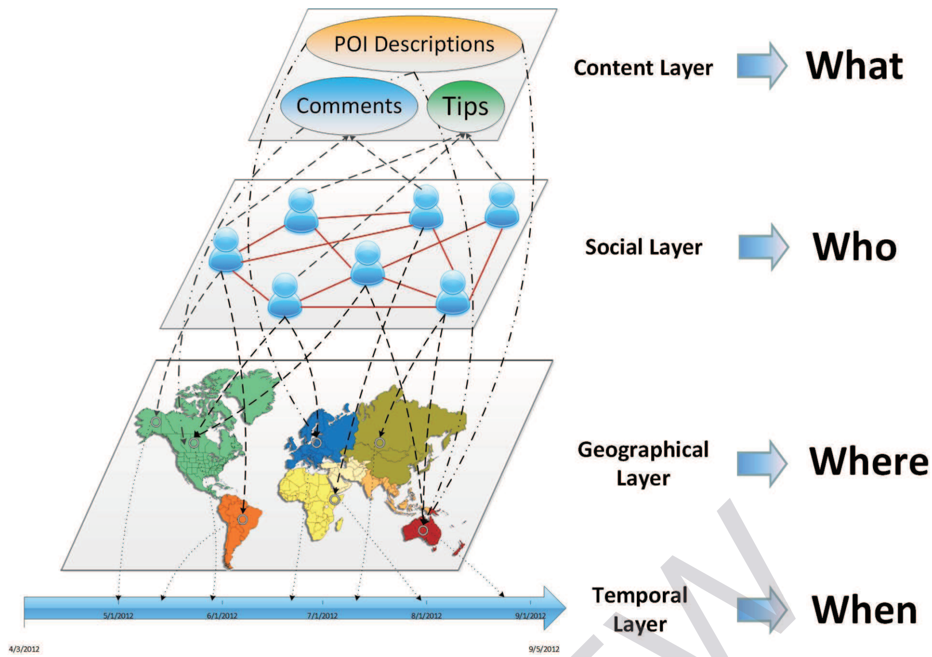

Figure 1.1: The Information Layout of Location-Based Social Networks

tual world, providing both opportunities and challenges for researchers to investigate users' check-in behavior for personalized POI recommendation in spatial ("where"), temporal ("when"), social ("who") and content ("what") aspects.

In the last decade, recommender systems have been widely studied among various categories, e.g., movie recommendation on NetFlix, dating recommendation on Zoosk, item recommendation on Amazon. However, it is not sufficient to directly apply these technologies as personalized POI recommendation on LBSNs presents unique challenges due to the heterogeneous information layout and the specificity of human mobility. Designing efficient POI recommendation approaches on LBSNs inevitably needs to consider the following properties.

1.1.1 Geographical Properties of Social Connections

Traditional social network analysis mainly studies network structure and properties without the consideration of geographical distance between nodes. Although the idea of "Death of Distance" proposed in 2011 claims that geographical distance plays a less important role due to the communication revolution and the rapid development of the Internet, which could make of our world a "global village" [7], studies on spatial structure of networks demonstrated that there is a strong correlation between geographical attributes and network properties, indicating the significance of considering the spatial properties of networks for future applications [25]. Researchers have further studied the distinctions between online and offline social networks [14], and discovered that geographical property does play important roles when constructing the social connection between two users especially in explaining their mobility in the physical world[67, 13].

As two special factors on LBSNs, Geographical property and social connections are coherent and affect each other in human behavior. For example, a user is more likely to be friends with other users who are geographically close to him, e.g, coworkers, colleagues. Likewise, a user may check-in at a location due to the influence from his friends, such as following friends' suggestions to visit a restaurant, going out with friends for shopping, etc. Such coherence results in a new property, commonly referred to as socio-spatial properties.

In recommender systems, user similarity evaluates how similar two users' preferences are, which is a significant measurement for recommendation especially social recommendation with collaborative filtering approaches. However, as discussed above, unlike regular social recommender systems, social connections on LBSNs exhibit unique geographical properties, providing a new dimension for computing user similarity. Therefore, considering the social information together with the geographical property enables us to capture the user preferences more precisely in POI recommendation on LBSNs.

1.1.2 Temporal Patterns of Geographical Check-ins

As suggested in [83, 12, 51], human geographical movement exhibits strong temporal patterns and is highly relevant to the location property. For example, a user regularly goes to a restaurant for lunch around 12:00 pm, watches movie on Friday night, and shops during weekends. This is generally referred to as temporal cyclic patterns. Such temporal patterns are not widely observed in other recommender systems. For instance, it is not common to observe a user regularly watching a specific movie (e.g., Batman, Avatar) or purchasing a specific item (e.g., camera, cellphone) at specific hour of the day, or day of the week. (Although birthdays or holidays like Thanksgiving may affect human behavior a bit, they are not commonly considered).

On the other hand, the temporal information of check-in actions on LBSNs is also considered as an order indicator to connect check-ins chronologically for generating location trajectories [98, 48, 81]. This is commonly referred to as temporal chronological patterns. For example, a user may want to sip a cup of coffee at Starbucks before he goes to office; or watch a move after dinner at a restaurant, and then relax at a bar.

In addition, temporal cyclic patterns and temporal chronological patterns are correlated to each other. Considering them together provides us a perspective to understand human mobility in terms of where a user would like to go at a specific time after his recent visits on other POIs. Thus, investigating the features embedded in temporal patterns enables us to better capture human check-in behavior, providing a potential opportunity to design more advanced POI recommender systems on LBSNs.

1.1.3 Semantic Indications of Check-in Content

Content information on LBSNs could be related to a user's check-in action, providing a unique opportunity for POI recommendation. When checking-in at a POI, a user may leave tips or comments to express his attitude towards the POI. Such content indicates abundant information w.r.t. the user's interested topics and personal preferences against various facets of the POI. For example, by observing a user's comment on a Mexican restaurant discussing its spicy food, we observe the *User Interests* in spicy food. If the comment is actually a compliment, e.g., "Best spicy food ever!", we could infer both the user's *Sentiment Indications* and her interests.

On the other hand, a POI is commonly associated with descriptive tags. Through studying these tags, one can not only infer the POI's property but also the interests of users who have checked-in at this POI. For example, by observing a POI's description as "vegetarian restaurant", we infer that the restaurant serves "vegetarian food" and users who check-in at this POI might be interested in the vegetarian diet. This is an example of *POI Properties*.

These three types of content information, i.e., *POI properties*, *User Interests*, and *Sentiment Indications*, are all related to a user's check-in actions and provide conceptual interpretations to three facets of his check-in actions, as listed in Table 1.1. In recommender systems, user interests and target properties are the two essential elements in capturing a user's action (e.g., check-in) on a target (e.g., POI) for recommendation [36], while user assessment has also been recognized as an important factor to gauge the check-in action for future recommendation [70]. Investigating them together makes it possible to infer how a user's interests match a POI's property and whether the user prefers to visit that POI. Thus, content information on LBSNs provides a conceptual perspective to investigate users' check-in behavior, which in

Table 1.1: Facets of Check-in Actions w.r.t. Content Information

Content Information	**Facets of Check-in Actions**
POI Properties	What is this POI about?
User Interests	Am I interested?
Sentiment Indications	How good is this POI?

turn constitutes the key factors of recommender systems, suggesting its potential for improving POI recommendation.

1.2 Problem Statement

Let $\mathbf{u} = \{u_1, u_2, ..., u_m\}$ be the set of users and $\mathbf{l} = \{l_1, l_2, ..., l_n\}$ be the set of POIs where m and n are the numbers of users and POIs, respectively. The problem of personalized POI recommendation on LBSNs is defined as:

Given a user $u \in \mathbf{u}$, a set of POIs (locations) $\mathbf{l}_u \in \mathbf{l}$ that u has checked-in, recommend him some POIs for his future visits based on the LBSN context (e.g., social connections, content information of check-ins, time stamps of check-ins) related to him, as illustrated in Figure 1.2.

For ease of presentation, we use POI, venue, and location as interchangeable terms in this dissertation. The recommendation algorithms discussed in this work are designed for individuals. However, they can be easily extended for group recommendation with aggregation strategies [91].

1.3 Contributions

The properties discussed above, i.e., geographical properties of social connections, temporal patterns of geographical check-ins, and semantic Indications of check-in

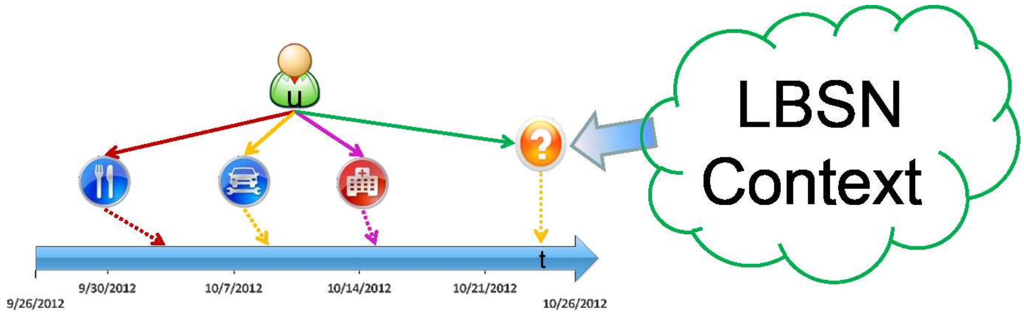

Figure 1.2: Illustration of Personalized POI Recommendation on LBSNs

Content, reveal the unique relationships of human behavior between geographical information and temporal, social, content information respectively, which are not commonly observed in other recommendation problems. In this dissertation, we study each property, and propose personalized POI recommender systems correspondingly, i.e., personalized geo-temporal recommendation, personalized geo-social recommendation, and personalized geo-content recommendation. To the best of our knowledge, this is the first work investigating these properties for POI recommendation on LBSNs. The contributions of our research are:

- Study the relationship between geographical check-ins and temporal information, model the temporal cyclic patterns and chronological patterns of a user's check-in behavior, and propose geo-temporal POI recommender systems regarding to these patterns with their complementary effect.

- Investigate the geo-social correlations of user check-in behavior to solve the "cold-start" POI recommendation problem and propose personalized geo-social POI recommender systems.

- Identify the challenges of analyzing semantic indications of content information on LBSNs, propose models to leverage such information for personalized geo-content POI recommendation.

1.4 Organization

The remainder of this dissertation is organized as follows. We first give a brief literature review in Chapter 2. From Chapter 3 to Chapter 5, we investigate the three LBSN properties to design personalized POI recommender systems. In Chapter 3, we introduce the personalized POI recommender system with geo-social correlations. We study the relationships between geographical distance and social friendships, and investigate them as a component w.r.t various facets. In chapter 4, we propose personalized geo-temporal POI recommender system. We study both temporal cyclic and temporal chronological patterns and their combinational effect. In Chapter 5, we analyze the user-generated content and POI-associated content, and leverage three types of content information including sentiment indications, user interests, and POI properties for personalized geo-content POI recommendation. We conclude the dissertation and point out promising research directions in Chapter 6.

Chapter 2

LITERATURE REVIEW

In the last decade, recommender systems have been widely studied among various categories, e.g., movie recommendation on NetFlix, job recommendation on Linkedin, item recommendation on Amazon, news recommendation on Yahoo. POI recommendation, also referred to as location recommendation, has been recognized as an essential task on recommender systems for enriching human life experience and facilitating decision making, which belongs to a sub-category of recommender systems. Thus, technologies of general recommender systems are also practically applicable to location recommendation, although the performance may be limited due to the specific properties of human mobility on LBSNs. In the following sections, we first give a literature review on general recommender systems, and then review techniques of location-based recommender systems for personalized POI recommendation.

2.1 General Recommender Systems

Recommender systems refer to technologies that help users find items of interest among a large amount of items by generating personalized recommendations [1]. The techniques of recommender systems can be generally classified into three categories: collaborative filtering, content-based, and hybrid models. Among them, collaborative filtering (CF) is one of the most successful approaches, which has been proven effective in practise [65, 71]. It requires a user-item rating matrix (i.e., user-location check-in frequency matrix) as an input. The fundamental assumption of CF is that if two users have similar behavior on the similar items (e.g., watching similar movies, buying similar products, visiting similar restaurants, etc.,), they will most likely have similar behavior on other items in the future. In general, collaborative filtering approaches can be further classified into memory-based CF and model-based CF. The memory-based CF approach leverages the entire user-item rating matrix for recommendation, which has been adopted in many commercial systems. According to whose similarity it relies on to perform the recommendation, approaches contain user-based [29] and item-based [65]. The idea of user-based CF is to capture a user u's preference on unvisited locations based on the preferences from K users most similar to him on locations. As an example of POI recommendation, it generally contains three steps:

1. Select K most similar users to u as his neighborhood $\mathcal{N}_u$.
2. Aggregate the preferences of users from $\mathcal{N}(u)$ on the locations not visited by u, deem them as u's preferences.
3. Rank u's preferences on those unvisited locations and select the top N locations for recommendation.

Analogously, item-based CF firstly finds K most similar locations and then calculates a weighted average of their check-in frequency.

Memory-based collaborative filtering approaches are efficient and easy to adopt. However, there are two shortfalls when it is applied to large-scale sparse data.

- **Sparsity**

  In many real-world applications, the user-item matrix is usually very sparse with a density of 10^{-4} to 10^{-5}. Under sparse data, the similarity measured from ratings (or check-in frequency) may not be reliable due to the insufficient information observed [60]. In an extreme case of "cold-start" problem, a new user who has no rating/check-in history would have the similarity value of 0 to any other users.