Expert Systems With Applications 86 (2017) 113–124

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Discovering socially important locations of social media users

Ahmet Sakir Dokuz[a,*], Mete Celik[b]

[a] Department of Computer Engineering, Omer Halisdemir University, 51245 Nigde, Turkey
[b] Department of Computer Engineering, Erciyes University, 38039 Kayseri, Turkey

ARTICLE INFO

Article history:
Received 6 January 2017
Revised 8 May 2017
Accepted 27 May 2017
Available online 29 May 2017

Keywords:
Socially important locations mining
Spatial social media mining
Historical social media data analysis
Social media networking sites
Twitter

ABSTRACT

Socially important locations are places that are frequently visited by social media users in their social media life. Discovering socially interesting, popular or important locations from a location based social network has recently become important for recommender systems, targeted advertisement applications, and urban planning, etc. However, discovering socially important locations from a social network is challenging due to the data size and variety, spatial and temporal dimensions of the datasets, the need for developing computationally efficient approaches, and the difficulty of modeling human behavior. In the literature, several studies are conducted for discovering socially important locations. However, majority of these studies focused on discovering locations without considering historical data of social media users. They focused on analysis of data of social groups without considering each user's preferences in these groups. In this study, we proposed a method and interest measures to discover socially important locations that consider historical user data and each user's (individual's) preferences. The proposed algorithm was compared with a naïve alternative using real-life Twitter dataset. The results showed that the proposed algorithm outperforms the naïve alternative.

1. Introduction

With the development of location-aware mobile devices, social media networking sites started collecting location information of their users when they are online. By using location data and social interactions (e.g., location-location interactions, user-location interactions, and user-user interactions), valuable information can be obtained, such as, which social friends are also real friends, which user likes what types of locations, which people spatially follow each other, which person's recommendation about a location gets more responses, which locations are important for a group of people, and so on. The aim of this study is to discover socially important locations that can be defined as places that are frequently visited by social media users in their social media life.

Discovering socially important locations is important for several application domains, including recommender systems, marketing (e.g., targeted advertisement applications), and urban planning. By discovering socially important locations, spatial preferences of a social group could be revealed.

However, discovery of socially important locations is challenging for several reasons. First, current interest measures are not sufficient to mine such locations. Second, the size and dimension of the data are growing over time. Third, human behaviors are unpredictable and can be affected by several factors.

Important locations mining problem is studied by many researchers and several solutions are proposed. However, these studies have several limitations. Some of these studies are based on GPS or Call Detail Record (CDR) data, thus proposed solutions are dependent to these factors. Another part of studies used social media datasets. Although several algorithms and frameworks are present for user-level recommendation and discovery tasks (Kefalas, Symeonidis, & Manolopoulos, 2016; Zheng, 2015), relatively small number of studies are performed for discovering group-level important/interesting locations of a group of social media users.

Example socially important locations for a social media user (twitter user) and a social media user group (1000 twitter users) can be seen in Fig. 1(a) and (b), respectively, in Istanbul, Turkey. Fig. 1 (a) shows three socially important locations of the social media user. The dataset used to discover the socially important locations of the social media user includes 1765 social media activities (i.e., location and time information of tweets) that belong to the period of December 2012 and December 2015. As can be seen in Fig. 1 (a), these locations are spatially unrelated locations and

* Corresponding author.
E-mail addresses: dokuz.a.sakir@gmail.com (A.S. Dokuz), mcelik@erciyes.edu.tr (M. Celik).

http://dx.doi.org/10.1016/j.eswa.2017.05.068
0957-4174/

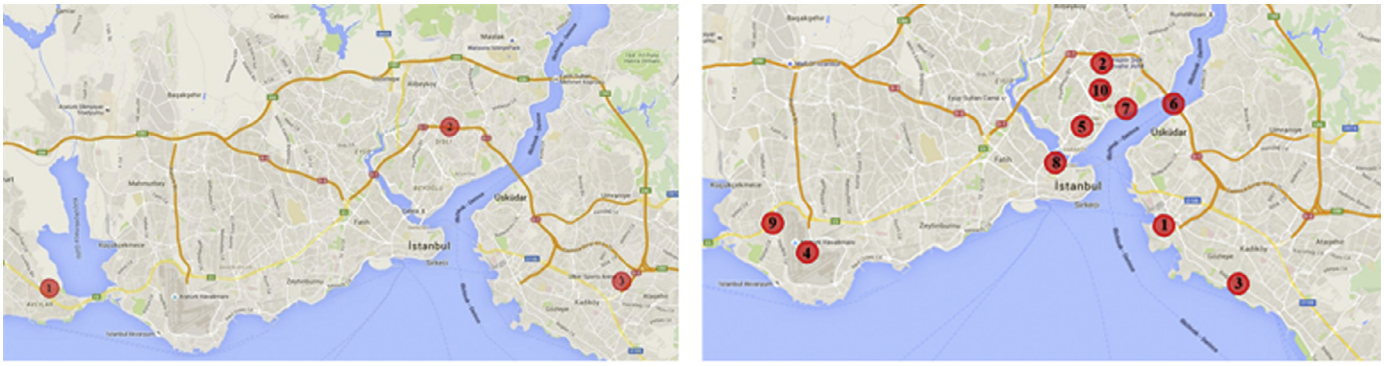

Fig. 1. Socially important locations of (a) a selected user and (b) a social media user group.

may have different meanings for the user. In location 1, there is a university campus, in location 2, there is a big shopping mall nearby, and in location 3, there is a big company's head office. We can make several predictions about importance of these locations for the user. For example, location 1 might be the user's university, location 2 might be a working place or shopping place of the user, and location 3 might be a working place. Fig. 1 (b) shows ten socially important locations of 1000 social media users in Istanbul, Turkey. The details of social media user group is presented in Section 5.2.2. Socially important locations of the group are clustered around the places where city social life is active. Comparison of the socially important locations of the social media user given in Fig. 1 (a) and that of social media user group given in Fig. 1 (b) shows that, the location 2, (i.e., the shopping mall) is a common socially important location for the social media user and the social media user group. In contrast, locations 1 and 3 are socially important locations only for the social media user not for the social media user group.

The three socially important locations of the user in Fig. 1 (a) are discovered based on his/her daily social media activity records. When social media users are brought together, the analysis of these users' datasets will lead to the discovery of important location(s) of that user group. For instance, if social media data of music lovers are analyzed, live music cafes, concert areas, and music shops are more likely to appear as socially important locations. If social media data of football fans are analyzed, then stadiums and sports-related locations are more likely to appear as socially important locations.

This paper deals with socially important locations mining problem over social media datasets. In this paper, we formulized interest measures of location density, visit lifetime and user prevalence to quantify socially important locations and proposed an algorithm, which is called SocioSpatially Important Locations Mining algorithm (SS-ILM), for discovering socially important locations efficiently. The proposed approach has two stages. In the first stage, user-level socially important locations are discovered using location density and visit lifetime interest measures. In the second stage, socially important locations of a social user group are discovered using user prevalence interest measure. The proposed algorithm, SS-ILM, was compared with a naïve algorithm. Both algorithms are experimentally evaluated using real dataset of Istanbul, Turkey Twitter users.

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 presents formal model of socially important locations mining problem. Section 4 introduces socially important locations mining algorithms, a naïve algoritm and proposed SS-ILM algorithm. Section 5 presents the experimental evaluation and Section 6 gives conclusions and future directions.

2. Related work

Spatial and spatio-temporal data mining have found a new research area after the introduction of Location Based Social Network (LBSN) datasets (Kefalas et al., 2016).

Spatial and spatio-temporal data mining is an important domain of data mining which aims to handle spatial and temporal data. Spatial co-location mining (Celik, 2015; Celik, Shekhar, Rogers, & Shine, 2008; Dadaser-Celik, Celik, & Dokuz, 2012; Yu, 2016) and spatial clustering (Hu & Sung, 2005; Tung, Hou, & Han, 2001) are some of the main topics (Shekhar, Lu, & Zhang, 2001). A co-location in social media datasets can be defined as a group of users or locations that are located neighboring spatial areas (Kefalas & Symeonidis, 2015; Vasuki, Natarajan, Lu, Savas, & Dhillon, 2011; Weiler, Schmid, Mamoulis, & Renz, 2015). The studies on mining classical spatial co-location mining assume that objects are boolean, that is, objects are present or not in a space or time. However, this assumption is not valid for social media datasets since a social media user may have more than one social media activity (e.g., sending tweets or posting messages on Facebook) at a location. Because of this reason, the algorithms and interest measures proposed in these studies may not be suitable to mine socially important locations out of social media datasets.

The important locations mining studies can be classified into two categories: 1) mining user-level socially important locations and 2) mining group-level socially important locations. Table 1 presents a summary of the literature related to this study.

The studies on mining user-level socially important locations include location/route recommendation, friend recommendation, activity/tag recommendation, interesting/important locations discovery, route mining, and discovering urban and mobility patterns (Table 1).

User-level location/route recommendation studies include the ones conducted by Bao, Zheng, and Mokbel (2012), Doytsher, Galon, and Kanza (2011), Kefalas and Symeonidis (2015), Levandoski, Sarwat, Eldawy, and Mokbel (2012), Liu and Seah (2015), Liu and Xiong (2013) and Ying, Lu, Kuo, and Tseng (2012). Liu and Seah (2015) proposed a method for important location (i.e., Point of Interest - POI) recommendation based on GPS trajectory data of mobile users and proposed density-based clustering approach to discover important locations. To determine the recommendation score of an important location (i.e., POI) three different factors (popularity, temporal and geographical features) have been used. Doytsher et al. (2011) presented a graph model that supports the representation of frequently-traveled routes in a socio-spatial network, proposed a query language to traverse the graph for querying social-based route recommendations, and used GPS trajectory dataset. In the studies conducted by Levandoski et al. (2012) and