



Contents lists available at ScienceDirect

Information Processing and Management

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

Discovering socially similar users in social media datasets based on their socially important locations

Mete Celik^a, Ahmet Sakir Dokuz^{*,b}

^a Department of Computer Engineering, Erciyes University, Kayseri 38039, Turkey

^b Department of Computer Engineering, Nigde Omer Halisdemir University, Nigde 51245, Turkey

ARTICLE INFO

Keywords:

Spatial social media mining
Socially similar users
Mining user similarity
Socially important locations
Twitter

ABSTRACT

Socially similar social media users can be defined as users whose frequently visited locations in their social media histories are similar. Discovering socially similar social media users is important for several applications, such as, community detection, friendship analysis, location recommendation, urban planning, and anomaly user and behavior detection. Discovering socially similar users is challenging due to dataset size and dimensions, spam behaviors of social media users, spatial and temporal aspects of social media datasets, and location sparseness in social media datasets. In the literature, several studies are conducted to discover similar social media users out of social media datasets using spatial and temporal information. However, most of these studies rely on trajectory pattern mining methods or take into account semantic information of social media datasets. Limited number of studies focus on discovering similar users based on their social media location histories. In this study, to discover socially similar users, frequently visited or socially important locations of social media users are taken into account instead of all locations that users visited. A new interest measure, which is based on Levenshtein distance, was proposed to quantify user similarity based on their socially important locations and two algorithms were developed using the proposed method and interest measure. The algorithms were experimentally evaluated on a real-life Twitter dataset. The results show that the proposed algorithms could successfully discover similar social media users based on their socially important locations.

1. Introduction

With the advances in smartphones, tablets, and other devices, social media networking sites (SNS) started collecting geographical information of users. Widespread use of SNS provides big spatial social media data for extracting valuable and interesting patterns. Spatial social media mining could provide modelling human behaviors in SNS and pattern discovery of users in social media datasets. Also, spatial aspect of the datasets could provide several information about correlations of social media users based on their common spatial behaviors. Recommending locations, measuring user influence, discovering common routes, discovering socially important locations, and mining socially similar users are some of the main applications of spatial social media mining (Bao, Zheng, Wilkie, & Mokbel, 2015; Kefalas, Symeonidis, & Manolopoulos, 2016; Riquelme & Gonzalez-Cantergiani, 2016).

This study focuses on discovering socially similar users based on their socially important locations. These locations are places that are frequently visited by social media users in their social media history (or lifetime) (Celik & Dokuz, 2017; Dokuz & Celik, 2017a; 2017b). For a social media user, a place, that is visited frequently and for a long time in his/her social media history, becomes his/her

* Corresponding author.

E-mail addresses: mcelik@erciyes.edu.tr (M. Celik), adokuz@ohu.edu.tr (A.S. Dokuz).

<https://doi.org/10.1016/j.ipm.2018.08.004>

Received 2 November 2017; Received in revised form 27 July 2018; Accepted 7 August 2018
0306-4573/ © 2018 Elsevier Ltd. All rights reserved.

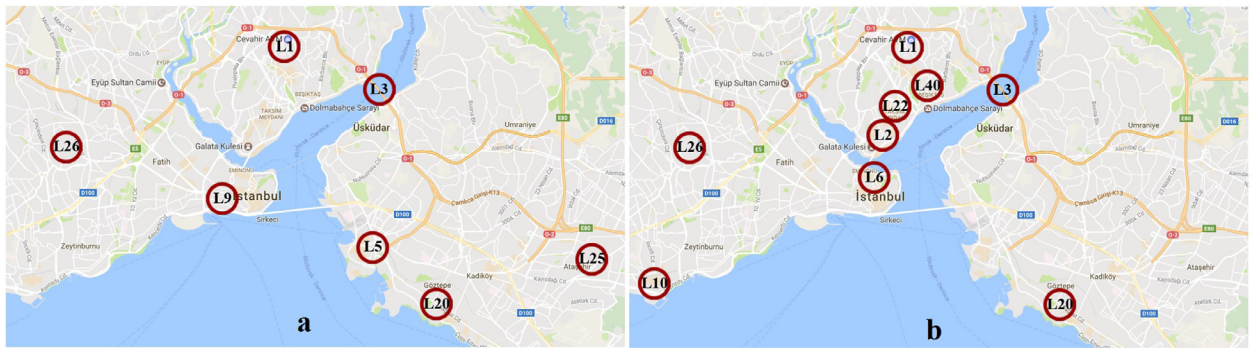


Fig. 1. Social important locations of two social media users (a) User 1, (b) User 2.

socially important location. Such places can be home, work, social place, etc. In this study, we take into account individual users' socially important locations to determine the similarity between the users.

Discovering socially similar users is an application of social media data mining that aims to discover similar users by analyzing their social media activities (Akcora, Carminati, & Ferrari, 2013; Cheng, Pang, & Zhang, 2015; Wan, Zhou, & Pei, 2017). It is beneficial for several application areas, such as, discovering communities in SNS, extracting common behaviors of social media users, detecting anomalous users, and mining similar interests of users within a geographic region (e.g., shopping center, city, country, etc.).

However, discovering socially similar users is challenging due to several reasons. First of all, social media datasets are huge and growing rapidly over time. Second, spam behaviors are so common in SNS and detecting spammers is challenging. Third, discovering socially similar users is computationally complex and requires computationally efficient algorithms. Fourth, user behaviors are hardly predictable. Fifth, geographical and temporal aspects of social media datasets bring extra load to algorithms. Also, location sparseness is another problem that hardens discovering socially similar users.

Several studies are available in the literature to discover similar users in social media datasets. These studies mainly focus on discovering similar users based on their trajectory, semantic, and location history information. However, these methods do not take into account socially important locations of users. They also do not take into account the temporal context (i.e. duration of visits and visits over time) of locations to discover socially similar users.

In this study, we propose algorithms and utilize Levenshtein distance to discover socially similar users based on their socially important locations. The proposed approach, first, discovers socially important locations of users, and then calculates the similarity of these users based on their socially important locations using Levenshtein distance metric.

Fig. 1(a) and (b) present socially important locations of two social media users. As can be seen in the figure, four locations are common for both users, such as, locations L1, L3, L20, and L26. Based on these locations of two users, we can observe that nearly half of their socially important locations are common. Also, the similarity in importance order of socially important locations increase similarity of the users.

1.1. Research objectives

The main research objective of this study is to discover similar social media users based on their socially important locations. Our specific research objectives are to examine the effect of using temporal context amongst other context types (i.e. trajectory, semantic, location history) and to use the Levenshtein distance type metric for ranking socially important locations. Socially important locations reveal spatial preferences of social media users, such as, which locations they like, which locations they frequently visit, etc. In the literature, trajectory, semantic, and location history based approaches have been used for discovering similar users. However, these methods do not take into account socially important locations of users and the temporal context (i.e., duration of visits and visits over time) of locations to discover similar users. In this study, we propose algorithms and utilize Levenshtein distance to discover socially similar users based on their socially important locations.

1.2. Contributions

The main contributions of this study are listed as follows:

- The definitions of socially similar users and discovering socially similar users using socially important locations are provided.
- A novel method of user similarity, which is based on socially important locations of social media users, is developed.
- A new interest measure, which is based on Levenshtein distance metric, is proposed to quantify user similarities.
- Two novel algorithms are proposed to discover socially similar users.
- Proposed algorithms are experimentally evaluated using real-life Twitter dataset and compared with a previous study.

The rest of this paper is organized as follows. Section 2 presents literature review. Section 3 presents the preliminaries of socially

Download English Version:

<https://daneshyari.com/en/article/10998000>

Download Persian Version:

<https://daneshyari.com/article/10998000>

[Daneshyari.com](https://daneshyari.com)