MinUS: Mining User Similarity with Trajectory Patterns

Xihui Chen[1⋆], Piotr Kordy[1], Ruipeng Lu[2], Jun Pang[1,2⋆⋆]

[1] Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg
[2] Faculty of Science, Technology and Communication, University of Luxembourg

Abstract. The development of positioning systems and wireless connectivity has made it possible to collect users' fine-grained movement data. This availability of movement data can be applied in a broad range of services. In this paper, we present a novel tool for calculating users' similarity based on their movements. This tool, MinUS, integrates the technologies of trajectory pattern mining with the state-of-the-art research on discovering user similarity with trajectory patterns. Specifically, with MinUS, we provide a platform to manage movement datasets, and construct and compare users' trajectory patterns. Tool users can compare results given by a series of user similarity metrics, which allows them to learn the importance and limitations of different similarity metrics and promotes studies in related areas, e.g., location privacy. Additionally, MinUS can also be used by researchers as a tool for preliminary process of movement data and parameter tuning in trajectory pattern mining.

1 Introduction

Due to the free access to GPS (global positioning system), people have access to their precise whereabouts. This access in turn leads to the collection of enormous amount of movement data, which offers us a new source of information to study human being's behaviour. For instance, we can check whether humans possess *swarm patterns* as birds and other animals during their movement. We can also learn the mutual influences between users' movements and their social relationships [1]. Among all interesting patterns, periodicity is naturally inherited in people's movement [2]. In other words, people have the intention to repeat some of their routes, possibly with similar temporal patterns. For instance, as a student of University of Luxembourg, Pierre has a daily routine from his residence to the train station where he takes bus to Campus Kirchberg. We call such repeated routines *trajectory patterns*. Intuitively, a trajectory pattern is a sequence of places of interests (PoIs) which a user frequently visits. With this interpretation, the daily routine of the student can be expressed as $residence \rightarrow train\ station \rightarrow Campus\ Kirchberg$. Typical transition time between two consecutive places of interest can also be added as part of a trajectory pattern. The extraction of trajectory patterns from users' travel history has been well studied in the literature and many mining algorithms have been proposed (e.g., [3]).

Trajectory patterns can be explored in many ways [4,5,6], one of which is friend recommendation in social networks. This is inspired by the fact that users' movement

---
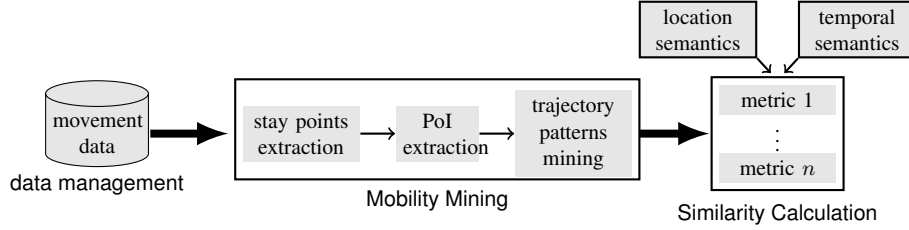
⋆⋆ To whom correspondence should be addressed.

Fig. 1. The architecture of MinUS.

reflects their interest. For instance, frequent visits to a reading club can indicate a user as a fan of literature. Thus, by the similarity between two users in terms of their trajectory patterns, we can infer whether they share similar hobbies. To the best of our knowledge, no tools are publicly available to calculate user similarity with trajectory patterns.

In this paper, we present a tool which integrates the technologies of trajectory pattern mining with recent research on user similarity calculation with trajectory patterns. Compared to existing trajectory pattern mining tools, our tool provides a graphical interface and allows users to control the mining process. More specifically, users have access to all intermediate results which can be visualised with our tool. Thus, researchers can use MinUS as a tool to preliminarily process movement data and tune parameters in trajectory pattern mining. With a series of metrics implemented, tool users can measure user similarity from different perspectives and thus learn the limitations and importance of different metrics. MinUS has one distinguishing feature that allows users to use the semantics of spatio-temporal information in trajectory patterns to mine user similarity.

Our tool is implemented with Java and C#. It is available online and can be downloaded from the following link `http://satoss.uni.lu/software/MinUS`.

2   The MinUS Tool

The MinUS tool has three function modules which are shown in Figure 1. The first module, *data management*, is in charge of managing movement datasets which are collected by different organisations. Such a dataset consists of a number of users whose movement is stored in the form of daily trajectories. This module keeps track of the statistic information about the users in each dataset. The statistics will be updated automatically once the values of the fields are available. The second function module, *mobility mining*, takes users' daily trajectories as input and outputs their trajectory patterns. The third module, *similarity calculation*, calculates the similarity values between users selected by tool users using the chosen similarity metric. In the following discussion, we give more details about the last two modules.

Mobility mining. In this function module, we implement the three sequential steps to calculate users' trajectory patterns according to the methodology given in [7,8]. At each step, a type of information about users' movement is mined. At the first step, we traverse all of the selected users' trajectories and detect their *stay points*, the centroids of small areas where a user stayed for a certain amount of time. A *place of interest* (PoI) is an area where users *frequently* visit, stay for a while and preform certain activities, such as

(a) MapView.   (b) User similarity calculation.
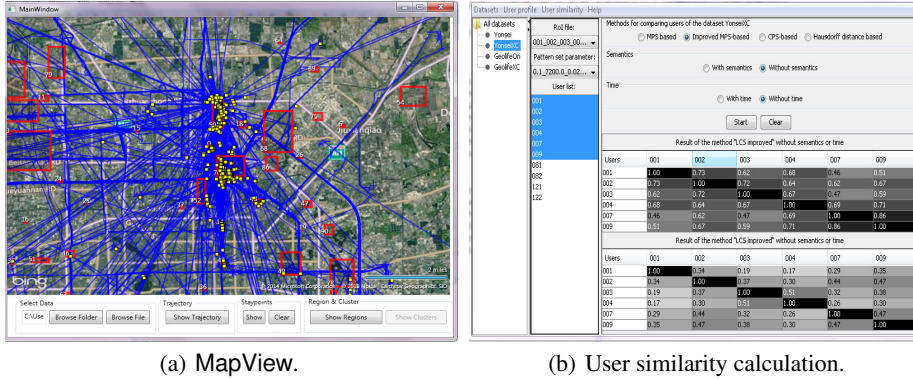
supermarkets and theatres. Based on this interpretation, at the second step, we calculate users' PoIs by identifying the regions where stay points are densely located. We implement a hierarchical clustering algorithm to calculate the clusters of stay points which are close to each other. Outlying stay points that are isolated from other points may increase the size of some PoIs. We make use of LOF (local outlier factor) to measure the extent of isolation of a stay point and remove a certain percentage of stay points that are most isolated. With extracted PoIs, users' trajectories are transformed into sequences of PoIs. At the last step, we explore the trajectory pattern mining tool of Giannotti et al. [3] to extract trajectory patterns.

Our tool gives tool users the control (by specifying the parameters required by the underlying algorithms) and the access to all intermediate results, such as users' stay points and PoIs. With a visualisation interface called MapView, tool users can put all the intermediate results on the map. In Figure 2(a), we show the screen shot of the visualisation interface. Users' trajectories are denoted by blue lines, while the yellow dots represent stay points and PoIs are depicted by red rectangles.

Similarity calculation. This module provides a platform to apply different metrics to measure user similarity with trajectory patterns. Figure 2(b) shows the interface of users similarity calculation. The tool users start with selecting a subset of users and the tool will return the similarity values between any two selected users. The results are visualised by a grid where the grey level of each cell indicates the similarity values between a pair of users. So far, we have implemented three categories of user similarity metrics: *maximal trajectory pattern based* [9,7,8], *common pattern set based* [10] and *Hausdorff distance based* [11]. Our tool also allows for taking into account *location semantics* and *temporal semantics* in the calculation. Location semantics denotes the functionalities of a PoI, e.g., restaurant and school, while temporal semantics represents the information revealed by time, e.g., weekends and weekdays. Since a PoI may correspond to a number of location semantics, we propose to use a probability distribution over all possible location semantics to demonstrate the uncertainty about a user's purpose in a PoI. Sometimes location semantics are used just for the purpose of comparison or validating the effectiveness of different metrics. For such cases, we implement a separate panel for tool users to simulate the distribution over location semantics for a PoI by giving the freedom to choose the number of location semantic tags and other factors.

The MinUS tool has been applied for experimental validation on different trajectory datasets [7,8,10]. We refer readers to those papers for the details.

3   Related Work

There have been many tools developed for mining users' movement patterns of different forms. The trajectory pattern miner by Giannotti et al. [3], used in MinUS, is developed to extract spatio-temporal frequent trajectory patterns from GPS-like trajectories of a set of move objects. Li et al. [4] design the tool MoveMine for scalable analysis on massive movement data, e.g., movement pattern mining and trajectory clustering. Two types of movement patterns are extracted: periodic patterns and swarm patterns.

By applying users' trajectory patterns, a few applications have been implemented. AllAboard by Berlingerio et al. [6] is a software system that makes use of cellphone data to improve existing public transport systems. Using users' call locations stored on cellphones, their trajectory patterns are extracted, with which new routes will be calculated and added to reduce users' waiting and travel time. Pelekis et al. [5] develop a tool called Hermoupolis to generate trajectories by simulating given trajectory patterns. This tool provides a method to synthesise trajectory datasets for researchers when real-life datasets are unavailable or not sufficiently large in experimental validation. So far, MinUS is the first tool which is publicly available to compare users' trajectory patterns.

References

1. Srivatsa, M., Hicks, M.: Deanonymizing mobility traces: using social network as a side-channel. In: Proc. CCS, ACM Press (2012) 628–637
2. Li, Z., Ding, B., Han, J., Kays, R., Nye, P.: Mining periodic behaviors for moving objects. In: Proc. KDD, ACM Press (2010) 1099–1108
3. Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D.: Trajectory pattern mining. In: Proc. KDD, ACM Press (2007) 330–339
4. Li, Z., Ji, M., Lee, J.G., Tang, L.A., Yu, Y., Han, J., Kays, R.: Movemine: mining moving object databases. In: Proc. SIGMOD, ACM Press (2010) 1203–1206
5. Pelekis, N., Ntrigkogias, C., Tampakis, P., Sideridis, S., Theodoridis, Y.: Hermoupolis: A trajectory generator for simulating generalized mobility patterns. In: Proc. ECML/PKDD. LNCS, Springer (2013) 659–662
6. Lorenzo, G.D., Sbodio, M.L., Calabrese, F., Berlingerio, M., Nair, R., Pinelli, F.: Allaboard: visual exploration of cellphone mobility data to optimise public transport. In: Proc. IUI, ACM Press (2014) 335–340
7. Chen, X., Pang, J., Xue, R.: Constructing and comparing user mobility profiles for location-based services. In: Proc. SAC, ACM Press (2013) 261–266
8. Chen, X., Pang, J., Xue, R.: Constructing and comparing user mobility profiles. ACM TWEB (2014) Accepted.
9. Ying, J.C., Lu, H.C., Lee, W.C., Weng, T.C., Tseng, S.: Mining user similarity from semantic trajectories. In: Proc. LBSN, ACM Press (2010) 19–26
10. Chen, X., Lv, R., Ma, X., Pang, J.: Measuring user similarity with trajectory patterns: Principles and new metrics. In: Proc. APWeb. LNCS, Springer (2014) Accepted.
11. Lu, R.: New user similarity measures based on mobility profiles. Master's thesis, University of Luxembourg (2013)