MovementFinder: A Multi-filter Visual Analytics Design for Movement Data Investigation

Siming Chen
Peking University

Zuchao Wang
Peking University

Zipeng Liu
Peking University

Zhenhuang Wang
Peking University

Chenglong Wang
Peking University

Zhengjie Miao
Peking University

Xiaoru Yuan*
Peking University

Abstract

In this paper, we propose MovementFinder, a multi-filter visual analytics system for movement data investigation. Our system integrates movement information from different datasets. Then with various visualizations and multiple filters, it is able to summarize the general movement patterns of a group of people, and help analysts detect abnormal events. Case studies demonstrate its effectiveness in a fictious analysis scenario.

1 Introduction

Movement data investigation is critical in visual analytics. The major tasks are often to summarize the general movement patterns and to extract abnormal movement events. As many important patterns and abnormalities are unexpected and non-obvious, visual analysis system have to be carefully designed to support such discovery.

In this paper, we propose MovementFinder, a multi-filter system to support data investigations from various aspects: location, time, people and event. Our system first combines related information from different datasets, including map, GPS tracks and transactions records. Then it visualizes above information with various views. Each view acts both as a visualization and a filter. Together they are able to support complex exploratory tasks.

2 Overview

Throughout this work, we use the fictious datasets from IEEE VAST Challenge 2014 Mini Challenge 2. All datasets are related to a country, Kronos. In the capital, Abila, a big company called GAStech experienced a kidnap. It is suspected that some employees assisted the kidnap, therefore the GPS logs of their cars are provided. The ownership of each car is recorded in a car assignment file. Besides, the transaction logs of the employees are provided, as well as a raster format tourist map of Abila, a vector format road network and a name list of the employees.

The GPS log and transaction datasets cover 54 employees and a time span of two weeks, Jan.6-19. There are 685,171 GPS records at 1 second resolution. There are two transaction datasets: a loyalty card dataset with 1,391 records, and a credit card dataset with 1,491 records. Most transactions are recorded in both datasets.

Our major tasks are as follows: 1) Describe the general daily life pattern of GAStech employees; 2) Detect the abnormal events or patterns among the employees. To make it concrete, we define the general patterns of employees according to the major types of locations they visit (e.g. office, restaurants), and the visit time (e.g. working hours, lunch hours). Further, we consider the following types of abnormal events: visiting a strange location, visiting a location at strange time, gathering of employees, large transactions.

To actually support above tasks, we follow the analysis pipeline proposed by Andrienko et al. [1, p.354]. We extract events from the movement data and then review them by looking back at the GPS tracks. Specifically, we view the movement as a sequence of stops at different locations [2], e.g. *home* $\rightarrow$ *restaurant* $\rightarrow$ *GAStech*.... We consider each stop as an event. Each event is associated with a person, a location, a time range and possibly a transaction. By directly plotting the GPS tracks and the event sequence for each person, we would have a basic idea of the general movement pattern. Then we visually detect potential abnormal events and review them with related GPS tracks and transactions. This usually involves filtering the movements of other people with similar titles, similar movement patterns, or at similar locations, or at the same time. Such comparison and correlation analysis help validate the potential abnormal events, and guess what is happening.

3 Preprocessing

In the preprocessing stage, we first detect events from the GPS logs. We define all stops above 1 minute in GPS logs as events. Each event naturally have a person, a time span and a location.

Then we enrich the event data with Point-of-Interests (POIs) and transaction information. For POI enrichment, we first manually extract the public POIs from the tourist map, and put them into many categories. Additionally, we try to identify the home of each employee as the most frequently visited location at 4:00 am. This is treated as a special POI. Then for each event, if its location is within the boundary of a POI, it would have a corresponding label, e.g. "GAStech", "restaurant", "shop", "home". Otherwise it would have a "non POI" label. For transaction enrichment, we first merge the credit card records and loyal card records. Then for each employee, if the transaction time is within the time span of an event, it is assigned to that event.

During the above preprocessing steps, we have to deal with uncertainties in various aspects:

POI data is manually extracted from tourist map. It is neither precise nor complete. Therefore we refined it later with GPS logs.

GPS data contains errors, e.g. a track can be systematically shifted. We manually corrected it. GPS data also have data missing: sometimes the distance between two consecutive records can be far away. We represented them as dotted lines in the interface, and consider them as special events with "uncertain" label.

Transaction data contains conflict in price and transaction time, because both credit card and loyalty card records are provided. In case of price conflicts, we show both prices. In case of transaction time conflicts, we select the time that better matches the GPS logs.

Car assignment data does not include the truck drivers. We manually determined the truck assignment by visually matching the GPS logs and transaction records.

4 Visual Interface

MovementFinder is a web-based application. As shown in Figure 1, it has five linked views, each acting as both a visualization and an interactive filter. It also has a control panel.

*e-mail: {csm, zuchao.wang, zipeng.liu, wangzhenhuang, wangchenglong, miaozhengjie, xiaoru.yuan}@pku.edu.cn

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

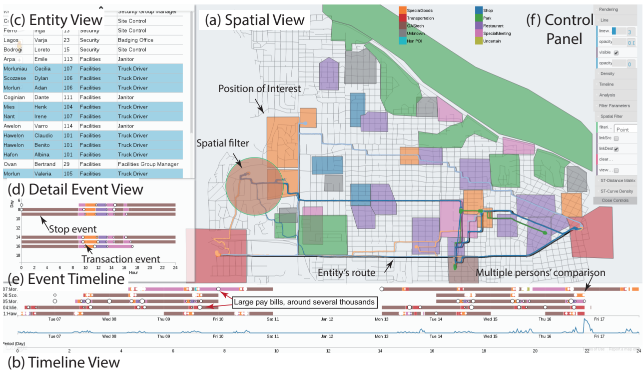

Figure 1: Interface of MovementFinder.

Map view shows the positions of POIs and GPS tracks (Figure 1a). Each POI is represented by a polygon, with color encoding the POI categories. Each GPS track is represented as a polyline. The events associated with the track can be highlighted, with their time mapped to the outer 24 hour ring (see Figure 3d). Users can apply spatial filters on the map to select GPS tracks passing single or multiple regions.

Timeline view shows the temporal distribution of GPS records (Figure 1b). Users can apply temporal filters on the timeline to select GPS tracks within single or multiple time ranges.

Entity view shows a name list of the employees (Figure 1c). Users can directly select people on the list.

Detail event view shows the whole event sequence of one employee (Figure 1d). Each row represents one day. The colored bars represent the time spans of events, with color indicating POI labels. If there are transactions associated with a event, a white circle will be shown. The size of the circle represents the price.

Event timeline shows the event subsequence in the selected time range for multiple employees (Figure 1e). It mainly used to compare/correlate the behaviors of different people.

Control panel is used for parameter adjustment (Figure 1f).

5 Case Studies

We use two cases to illustrate the functions of the MovementFinder. The first one is about general pattern investigation and the second one is about abnormal event detection.

5.1 General pattern investigation

Generally, we can see three peaks very weekdays on the timeline (Figure 2a). Taking one employee Azada Lars for example, we check his event sequence in the detail event view (Figure 2c). Notice the brown band represents GAStech, gray for home and blue for restaurants. Then we focus on one day, Jan. 10th, and the map view shows his GPS track on that day (Figure 2b). With temporal filters, we can check different parts of the track separately (Figure 2d). The time of event on this GPS track is mapped to the outer 24 hour ring.

5.2 Abnormal event detection

Initially, we found there were some small amounts of traffic at midnight of Jan. 9th. Therefore we select these unusual GPS tracks with temporal filters. Among the tracks, we found there's an intersection near the Taxiarchon Park (highlighted in green in Figure 3). Two people, Mies and Bordrogi are involved, and their event subsequences in this period are shows in the event timeline (Figure 3d). Mies went to the park at 23:06 and left at 3:30 on the next day (Figure 3b). Just after Mies left, Bordrogi arrived at 3:32 and stayed there till 7:23 (Figure 3c). Then Bordrogi had breakfast in a cafe different from the one he usually went to, and he didn't pay for bill.

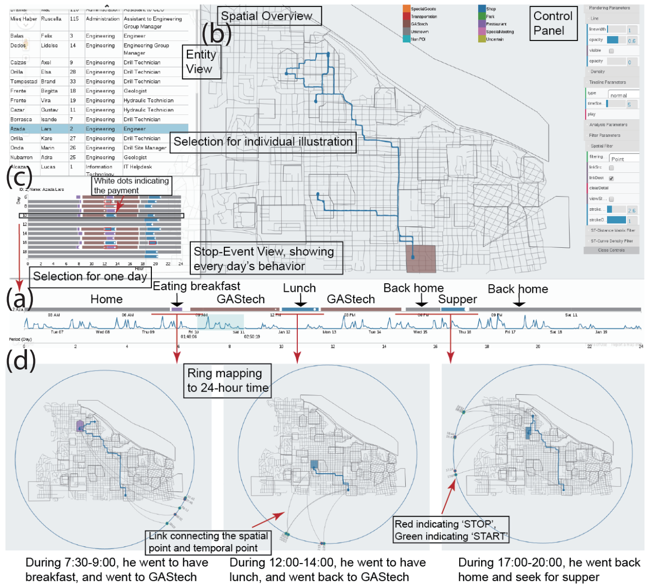

Figure 2: General life pattern of a typical GAStech employee.

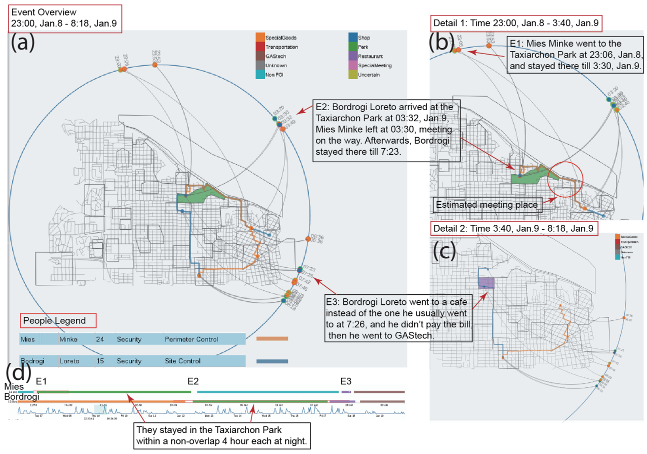

Figure 3: Abnormal behaviors of two employees near Taxiarchon Park at midnight.

This is suspicious, because it indicates there might be someone else with him and paid the bill.

6 Conclusion

In this paper, we have presented a visual analytics system for movement data investigation. Our research emphasizes the data preprocessing and a multi-filter visual design. In the future, we would try to handle larger scale of data. Besides, we would provide a collaborative visual analytics environment to solve complex problems.

Acknowledgements

The authors wish to thank IEEE VAST Challenge committee and reviewers. This work is supported by NSFC No.61170204

References

[1] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual Analytics of Movement*. Springer, 2013.

[2] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot. A conceptual view on trajectories. *Data Knowl. Eng.*, 65(1):126–146, 2008.