VAST 2014: Summary on Mini Challenge Two

Carlos K.F. TSE
ASTRI
Hong Kong
carlostse@astri.org

Yang LIU
ASTRI
Hong Kong
yangliu@astri.org

Abstract

VAST 2014 mini challenge two required us to analyze movements and tracking data in a fictional company employee missing event. This paper includes our design and software to tackle the challenge. In addition, the procedure to find the abnormal patterns in the given data is illustrated.

Keywords: VAST 2014, mini challenge two, visual analysis.

Index Terms: [Information systems]: Information retrieval-Retrieval tasks and goals-Information extraction; [Computing methodologies]: Artificial intelligence-Computer vision-Computer vision problems-Matching; [Information systems]: Information systems applications-Decision support systems-Data analytics

1 Introduction

In the story given by VAST 2014, the Tethys-based gas company GAStech has made huge profits in the island country of Kronos. However, many Kronos people suffered from pollution due to the GAStech commercial activities. On January 20th 2014, several employees of GAStech were missing. Protectors of Kronos (POK), one local environment protection organization seemed to be the suspect behind the scene. Accordingly, three mini challenges are presented as well as one grant challenge to help further investigations. Specifically, mini challenge two provides a data set of over six hundred thousand GPS records, thousands of credit card transactions and loyalty cards records in the two weeks before January 19. The focus there is the abnormal patterns derived from all the data available. [1]

In order to find the abnormal patterns required, the normal patterns should be firstly determined. There are two major issues found in the analysis, they are how to match the GPS data to various locations, such as restaurants, shops, companies and apartments, and the other is how to define a proper pattern which will be useful for investigation.

The rest of the paper is organized as follows. In sections two, the software used and developed are introduced. Specifically, how to match the GPS data to location is clarified. In section three, the analysis procedure and results are presented. Finally, we summarized our findings.

2 Software

We build a database to deal with the GPS data, credit card transactions and loyalty cards records by extracting the information from raw data. The database is stored in SQLite. The GPS data only contains the date time, latitude and longitude. Since Kronos is not a real place, we cannot use the existing GPS tools in the market, such as Google Map API to determine the locations. In addition, the loyalty card and credit card records contain only date, that is, no time is given. Accordingly, we have to link up the credit card and loyalty card transaction records with the GPS data with following assumptions. First, we assume the employees drive their own cars for shopping, meals, and work. Secondly, we assume after they park the car with a reasonable amount of time, for example 15 minutes, they will use their credit cards and loyalty cards. Finally, we assume employees should stay at their apartments in the nights. An example of the location concluded is shown in Figure 1.

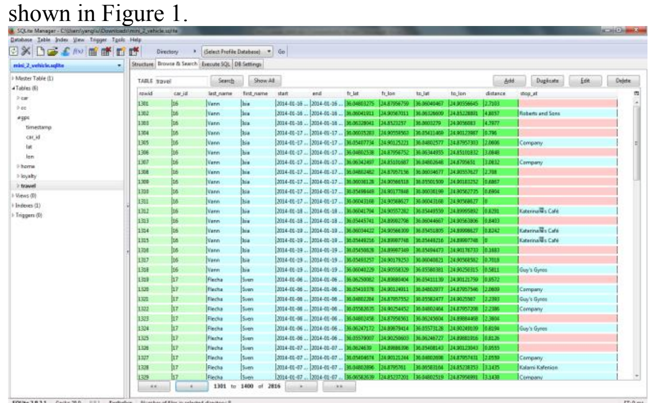

Figure 1: Location Concluded for Isia Vann in SQLite.

After finding the locations of the restaurants, shops, offices and apartments, we analyze the daily patterns of the employees in two ways. One is to draw the latitude and longitude on the white background for all routes with the same scale with Python imaging tools (Figure 2), and the other is to classify the locations and draw the patterns in the line chart by HTML5 (Figure 3).

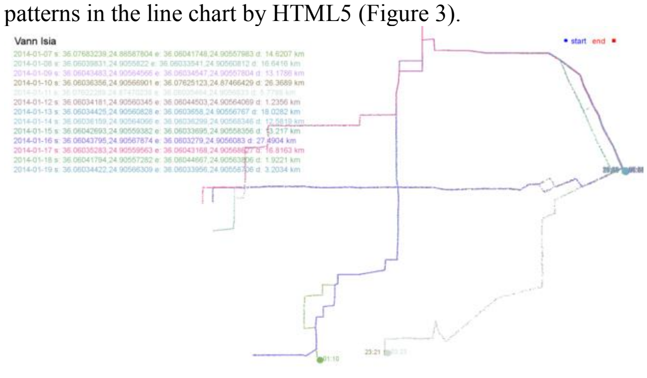

Figure 2: Route Traffic Pattern Example for Isia Vann.

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France

Figure 3: Line Chart Traffic Pattern Example for Isia Vann.

The route traffic pattern is easy to track. For example, if two employees have overlapped routes in a same interval, they probably meet each other around the given time. As a contrast, the line chart traffic pattern helps us to identify normal daily pattern quickly, as well as the abnormal traffic patterns.

Also, the consumption records can be visualized in the view of consumer and/or location. Therefore, it is convenient to identify multiple patterns for further investigation by checking the figures generated.

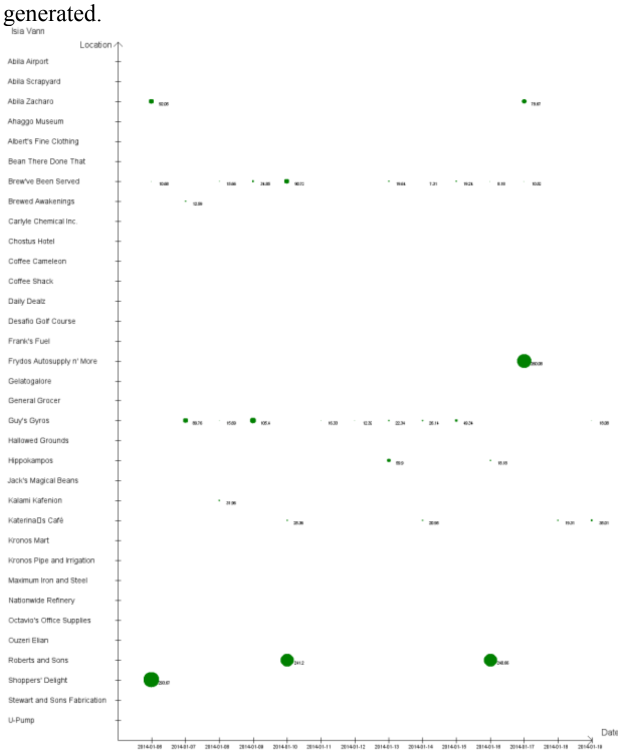

Figure 4: Consumption patterns for Isia Vann.

3 Discussion

3.1 Traffic Pattern Analysis

We start with employees with abnormal traffic patterns since we quickly found some employees go out at midnight multiple times or only once. For exmpale, Isak Baza, Linnea Bergen, Nils Calixito, Axel Calzas, Lidelse Dedos Adra Nubarron, Brand Tempestad had single late out Jan.10 middle night. But, since logically several meetings are needed for plan the people missing event, we consider multiple late outs be more important. In addition, although Lucas_Alcazar has multiple late outs, the GPS data shows he always stays at Gastech caompany, which indicates he is just working over time. Therefore, the final suspect list includes Isia Vann, Loreto Bodrogi, Hennie Osvaldo and Minke Mies.

To confirm the employees with multiple going out in the midnight are really having meetings, we write a distance calculator program. The input is the employees and the time, and output is the corresponding distance between them. Limited by the discrete GPS data, the calculator can only find out the nearest location of the input time and calculate the distance between them. The calculator tells us that some employees did have meetings in the mid-night, as the example shown in Figure 5.

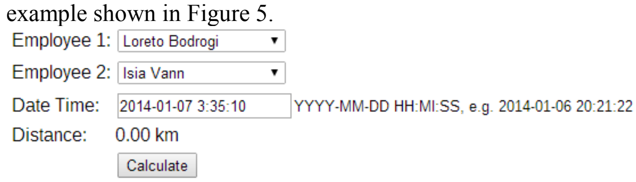

Figure 5: Distance Calculated Example.

Accordingly, the places visited in the middle night frequently by the suspects are plotted in Figure 6. The identified places are also matched to the tourist map given in the data set for further investigation.

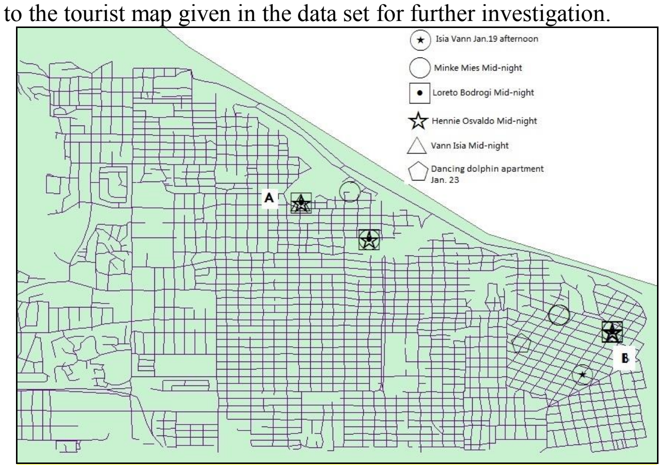

Figure 6: Location Analysis Example.

3.2 Consumption Pattern Analysis Results

First, we checked the difference of credit cards and loyalty cards per location. However, the difference is not evident. Therefore, we use the combined records only for simplicity. Secondly, with the suspects found in traffic patterns, we study the corresponding consumption patterns. For example, in Figure 4, we see Isia Vann has an abnormal large bill at Frydos Autosupply n'More at Jan. 17. But since he had a company car, the pattern does not clearly indicate he is a suspect in the kidnapping. Finally, the consumption record charts by location may reveal people gathering information if we assume they are sharing the bill. But, again, it is hard to tell it is a normal office gathering or kidnapping plan meeting. Accordingly, we do not include much consumption pattern analysis result in the final submission.

4 Conclusion

To the help of our visual analysis tools and analysis, we define some patterns for VAST 2014, mini challenge two. Accordingly, we found normal patterns as well as abnormal patterns for further investigation. The most useful results we found are the traffic pattern analyzed and the related suspect locations.

References

[1] *Vast Challenge 2014*, from http://vacommunity.org/VAST+Challenge+2014
[2] *Python Imaging Library (PIL)*, from http://www.pythonware.com/products/pil/
[3] *CanvasJS*, from http://canvasjs.com/