Saptiotemporal Analysis of Gastech Employees' Movement Data

Chen Yuxi
Singapore Management University
yuxi.chen.2020@mitb.smu.edu.sg

Lim Yong Kai
Singapore Management University
yongkai.lim.2020@mitb.smu.edu.sg

Jovinka Hartanto
Singapore Management University
jovinkah.2020@mitb.smu.edu.sg

ABSTRACT

The research paper should include an abstract of not more than 300 words. The actual research paper itself should not more than 6 pages excluding figures, tables, formula and references. The practice research paper must be edited by using R Markdown and the ACM: Association for Computing Machinery template of rticles should be used.

1. INTRODUCTION

The VAST 2021 Mini Challenge 2 outlines a hypothetical scenario in which several GAStech employees have gone missing and the organisation, Protectors of Kronos (POK), is suspected of being involved. The dataset includes two weeks of GPS tracking data for company cars assigned to employees, credit and loyalty card transactions of employees before the disappearance. ESRI shapefiles for the city of Abila & country Kronos are also provided. The challenge requires identification of suspicious activities hidden in data and determine any dubious people and locations that should be reported to the law enforcement.

The dataset was wrangled to develop a Shiny app that aims to provide users actionable insights based on the following analyses:

- Exploratory Data Analysis (EDA) of GPS tracking data and credit and loyalty card transaction patterns
- Visualisation of employee movements over time and associated purchase transactions
- Network analysis of employees based on their visited locations

This paper details our efforts to design, create and implement a web-based analytics tool to assist users from the law enforcement to derive insights and accelerate the investigation process of the disappearance of GAStech employees. The paper consists of six sections. Section 1 presents a general introduction of the paper. Section 2 provides an overview of the motivation and objectives of our project. Section 3 provides a review of techniques used in Vast Challenge 2014 and similar spatiotemporal visualisations. This is followed by a detailed description of the design principles used and data visualisation elements built, and a demonstration of the user interface in section 4 and 5. Insights and uses of the system are documented in section 5. Finally, the paper concludes by highlighting how the system can be extended or refined in the future.

2. MOTIVATION AND OBJECTIVES

Our research and development effort were motivated by the general lack of effective and easy to use web-enabled data visualisation tool to conduct data analysis on the GAStech employees related dataset. The project aims to enable the Kronos law enforcement the ability to easily analyse, drill down and identify key suspects and suspicious locations, and thereby speeding up the investigation process.

The use cases of the data visualisation tool include but are not limited to the following:

- The most popular locations and when they are popular
- Infer the owner of each credit card and loyalty card
- Identify potential informal or unofficial relationships among GASTech personnel
- Analyze unusual and suspicious activities of the employees
- Analyze daily routines of GasTech employees

3. REVIEW AND CRITIC OF PAST WORKS

The project drew inspirations from previous works by Singapore Management University (SMU) students and submissions of the 2014 Vast Challenge.

In the project, Dino Holmes Series, (https://wiki.smu.edu.sg/1617t1ISSS608g1/ISSS608_2016-17_T1_Assign3_Ong_Han_Ying), heatmap enabled readers to easily determine patterns and trends over a time. However, static heatmap are not reader friendly enough to determine what was the count at specific time slots. As each heatmap box represents a discrete count by using a gradient color fill, it was difficult to accurately determine the specific count. The analysis may benefit from making the heatmap plot interactive, allowing

the details to be displayed at the tooltip when hover across. This would allow granularity data to be more well-presented in the report.

Submission from Fraunhofer IAIS and City University London (Andrienko, Andrienko, and Fuchs 2014) and RBEI-Bangalore (Singhal et al. 2014) both used network cluster and analysis to investigate the relationships between GAStech employees. Fraunhofer IAIS university used an ego-centric graph whereas RBEI used a combination of fragmented and node-only layout to visual the relationship by connecting employees. Network analysis is an informative visualisation that provides an overview of potential relationships between employees or even connecting employees to different mediums such as the locations or emails.

Although network analysis provides an overview of the relationships between nodes, usually the plot will be cluttered which make it difficult to drill down to specific or individual relationships. An alternative would be to make the plot interactive so that readers will be able to drill down on specific areas to investigate the relationships.

4. DESIGN FRAMEWORK

A detail description of the design principles used and data visualisation elements built (Refer to Section IV: Interface of this paper.

4.1 Data Preparation

All the data wrangling and cleaning were performed using R.

4.2 Exploratory Data Analysis

5. DEMONSTRATION

6. DISCUSSION

What has the audience learned from your work? What new insights or practices has your system enabled? A full blown user study is not expected, but informal observations of use that help evaluate your system are encouraged.

7. FUTURE WORK

A description of how your system could be extended or refined.

References

[1] Fenner, M. 2012. One-click science marketing. Nature Materials. 11, 4 (Mar. 2012), 261–263.

[2] Meier, R. 2012. Professinal Android 4 Application Development. John Wiley & Sons, Inc.