Making Sense of Daily Life Data: From Commonalities To Anomalies

VAST 2014 Mini Challenge #2

Ji Wang*
Virginia Tech

Peng Mi†
Virginia Tech

Chris North‡
Virginia Tech

ABSTRACT

In this paper, we report the approach and results on the VAST 2014 Mini-Challenge 2: Analysis Movement and Tracking data of GAStech Employees' daily lives. Based on the commercial interactive visualization software Tableau[1], we follow the sense-making loop for analysis of the massive multi-dimensional, multi-source and time-varying data sets. The findings show that we can effectively identify the patterns and discovery the anomaly from these complex data sets.

Keywords: visual analysis, Geo-visualization, sense-making loop, intelligence analysis, human information interaction

Index Terms: H.1.2 [Use/Machine Systems]: Visual Analytic K.5.2 [User Interfaces]: Graphical user interfaces—Information Visualization

1 INTRODUCTION

In VAST Challenge 2014 Mini Challenge 2, participants work as an expert in visual analytics to make sense of the GAStech employees' daily life patterns and identify the suspicious behaviors which may be related to the missing staff members. The related data sets include two weeks GPS tracking data, car assignment records, map materials of Abila and Kronos, loyalty and credit card transactions. Besides, the data are imperfect, featuring with "uncertainties, with missing, conflicting data".

To investigate the knowledge behind these messy data sets, our first idea is to develop a powerful visual analysis software. Facing with the multiple sources and types of data, we found that it was impossible to develop an uniform tool for handling the images, texts and GPS tracking data within several weeks. Instead, we try to seek the commercial visualization software and aim to effectively leverage the existing tools for the problem solving. So as an visual analysis expert, we focus on the analytical reasoning processes for efficiently revealing the knowledge behind the data sets. We follow the sense-making loop (Figure 1, Pirolli & Card [2]), which guides our analysis based on Tableau visualization software, to discover the patterns and anomalies among the dirty data sets. In the following paper, we will describe our findings based on the sense-making loop, especially its two sub-loops: foraging of interested or relevant information and sense-making for hypothesis and schematizing.

2 VISUAL ANALYSIS

2.1 Commonalities Identification

To investigate the GAStech employees' daily routines, we first visualize the GPS tracking data by the scatter plot matrix, as shown in Figure 2. We find that most of the employees have regular drive routines during the weekdays. We also found some of GPS devices

---

*e-mail: wji@vt.edu
†e-mail:mipeng@vt.edu
‡e-mail:north@cs.vt.edu
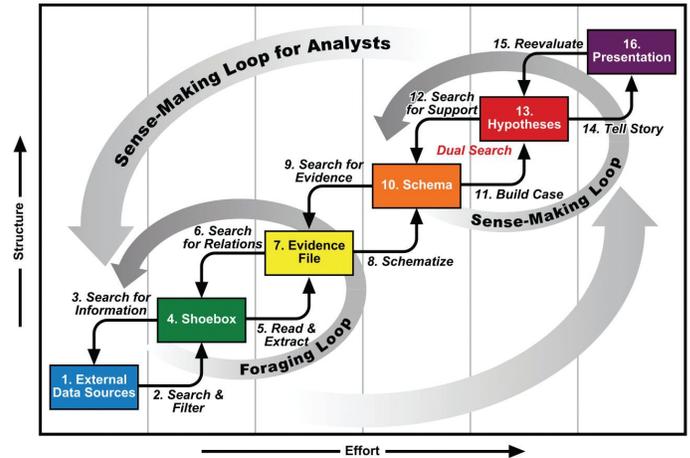
Figure 1: Sense-Making Loop

did not work well, and feature missing data or large noises. The last two rows of Figure2 show the truck routines, which they always cover the GASTech office and airport. Figure 3 also inspire us to hypothesise that most employees have regular schedule based on the temporal heatmap of cars' activities. The visualization of credit card also verifies it by the temporal heatmap in Figure 4. The employees are usually go to coffee shops during 7am - 8am in the morning and drive out for their lunches at noon with regular patterns.
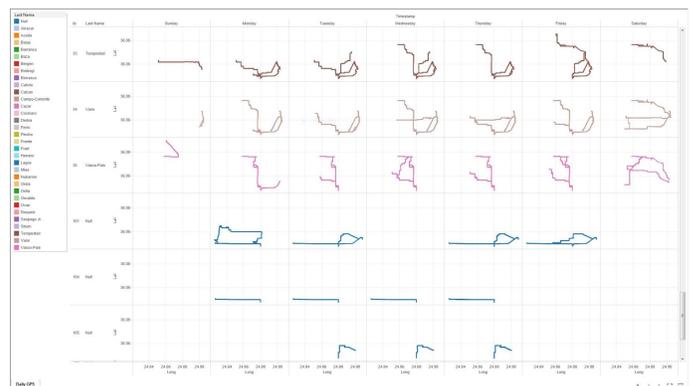
Figure 2: Scatter plot matrix of GPS latitude and longitude

The above analytical reasoning procedure is a typical sense-making loop for hypothesis generation and verification, which efficiently guides our discovery for the GAStech employees' daily routine.

Figure 3: Temporal Heatmap of the GPS activities
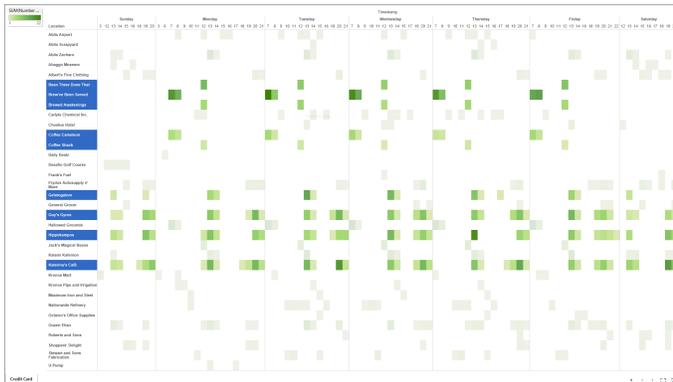
Figure 4: Temporal Heatmap of credit card transactions

2.2 Anomalies Detection

To further detect the hidden anomalies, we filter the GPS data by focusing on the employees' activities off their work. By synthesizing the GPS data with car assignment records, as well as the map materials of Abila and Kronos, we want to relate the 3Ws:"Who","When" and "Where".
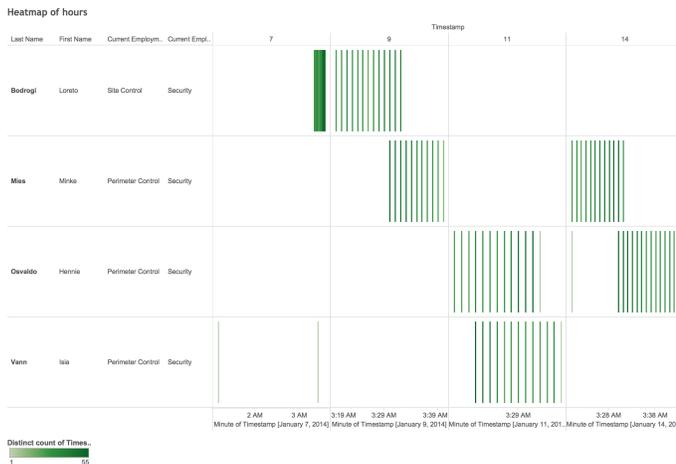
Figure 5: The time-line visualization about "Who" and "When"
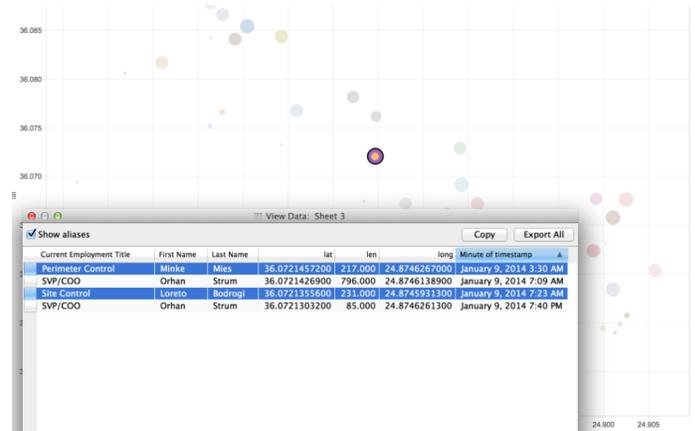
Figure 6: The scatter plot visualization about "Who" and "Where"

Figure 5 describes the relationship of "Who" and "When". Only four GAStech security staff have activities during 2am - 4am. Besides, this figure shows that every two staff have activities at the exact same time of each day's midnight. Figure 6 demonstrates the connections between "Who" and "Where" on Jan 14, 2014 around 12pm. We find that two security staff and GAStech SVP/CFO appear at the same places. By foraging more details of their activities, we understand that these two security staff stayed near the GAStech SVP/CFO house in shifts. In all, we forage the related data based on 3Ws, and speculate that four GAStech security staff (Minke Mies, Loreto Bodrogi, Isia Vann and Hennie Osvaldo) carried out the surveillance on executives' houses in pairs. These unusual events happened about Ada Campo-Corrente (SVP/CIO) on January 7th, Orhan Strum (SVP/COO) on January 9th, Willem Vasco-Pais (Environmental Safety Adviser) on January 11th, and Ingrid Barranco (SVP/CFO) on January 14th.

Our 3Ws' foraging loop effectively synthesizes the relevant data for discovering the unexpected events, linking the evidences and assumptions based on the Tableau visualization and interaction.

3 CONCLUSION

This paper shows our analytical reasoning processes about GAStech employees' daily routine based on the commercial visualization software Tableau. Our results show that we can effectively find the important patterns and anomalies by leveraging the existing tools. While our solution can not discover all possibilities, we demonstrate that combing sense making loop with existing softwares is an efficiently way for problem solving, especially the time-critical tasks.

REFERENCES

[1] J. M. Pat Hanrahan, Chris Stolte. Visual analysis for everyone. 2007.
[2] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. pages 2–4, 2005.