Safeguarding Abila: Spatio-Temporal Activity Modeling
VAST 2014 Mini Challenge 2 Award: Honorable Mention for Effective Presentation

Parang Saraf*   Patrick Butler†   Naren Ramakrishnan‡

Discovery Analytics Center
Department of Computer Science
Virginia Tech

ABSTRACT

We introduce a system for visual analysis of GPS tracking and financial data. This system was developed in response to VAST Mini-Challenge 2 and comprises of different interfaces for mining spatio-temporal and financial data.

Index Terms: H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces—Interaction styles (e.g., commands, menus, forms, direct manipulation)

1 INTRODUCTION AND PROBLEM OVERVIEW

The VAST 2014 Mini Challenge 2 describes a hypothetical scenario where some of the employees of an imaginary organization, GAStech have gone missing and it is speculated that an environmental activist group, Protectors of Kronos (POK) is responsible behind the disappearance. The provided dataset includes two weeks of GPS tracking data for several company cars assigned to employees, credit & loyalty card information of employees and ESRI shapefiles for the fictional city of Abila & country Kronos. The challenge requires identification of suspicious activities hidden in data and prioritization of all such activities in order to determine any unscrupulous persons and locations that are worthy of reporting to law enforcement agencies.

2 SYSTEM DESIGN

We developed a web-based visual analytics system for analyzing geo-spatial, temporal and financial transaction data. The system provides several widgets that empower an analyst to unearth aberrations buried in data. Google Maps was used to visualize spatial data and the Javascript-based graphical libraries d3.js [1] & nvd3.js were used for plots.

2.1 Geo-Spatial Data Analyzer

The creators for Mini Challenge 2, quite strategically, left out *point type* data in the provided ESRI shapefiles. Absence of this crucial information makes it difficult to identify "Points of Interest (POI)". A POI is defined as a specific point location that someone may find useful or interesting, e.g., shopping center, restaurant, etc. In order to identify POIs by using only the provided GPS tracking data, following methodology is implemented:

1. For each of the cars, all the geo-coordinates where the car is stationary for more than 5 minutes are identified. These points refer to places that users might have found useful.

2. Spatial clustering is performed on these geo-coordinates in order to represent points in the spatial vicinity of each other as

---

*e-mail: parang@cs.vt.edu
†e-mail: pabutler@vt.edu
‡e-mail: naren@cs.vt.edu

one single location. These clusters have a radius of 25 meters and describe a POI which is frequented by users.

3. Establishment names such as *xyz coffee shop* are associated with these spatially identified POIs. The step involves characterization of POIs as home, work and recreational locations. The system provides three different interfaces for characterizing POIs.

The interface for characterizing recreational POIs (see Figure 1), utilizes credit card swipe time information along with user location to visualize a recreational establishment on map. It is assumed that if a user is swiping his card at a particular establishment, then he is present there. Based on this assumption, credit card transactions are grouped by recreational establishments. Selecting a particular establishment displays the location of all the customers that visited that establishment in the past two weeks. The interface also displays credit and loyalty card transaction information in tabular format for each of the establishments.
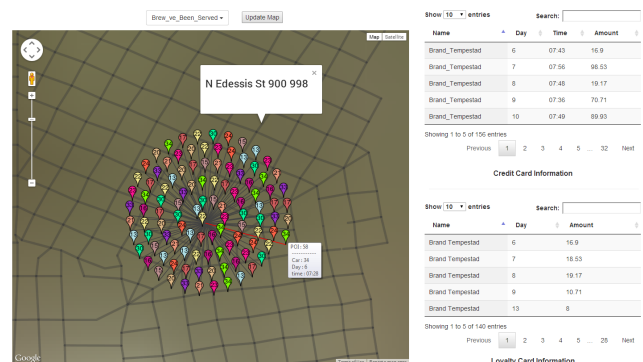

Figure 1: Characterizing Recreational POIs

The second interface provides a more generic platform for analyzing all types of POIs. It introduces two analytical views, viz. POI distribution over time and POI frequency over time. POI distribution over time (see Figure 3) is a scatter plot that displays all the POIs where a particular user was present over the 2 week duration. This helps in identifying home and work POI for each of the users by assuming that a user spends his nights at home and his working hours in the office. The widget also allows comparison of POI distribution for several users.

POI frequency over time (see Figure 4) displays the total number of users present at a particular POI during the 24 hour window. The plot helps in characterizing POIs as home, recreational, or work locations. For example, if several users are present at a particular POI during work hours (8 am till noon and 2 pm till 5 pm), then that POI would be classified as an office building.

The third interface (see Figure 5) allows for query and visualization of spatio-temporal GPS data. Using this interface, an analyst can generate selective playback of locations sequences visited
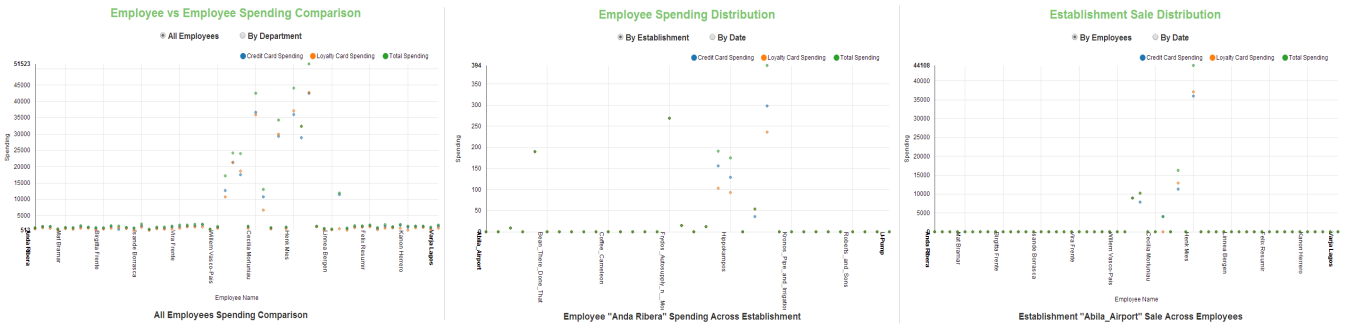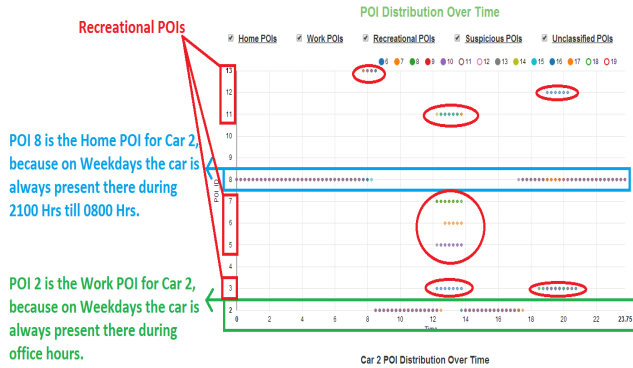
Figure 2: User Spending Comparison

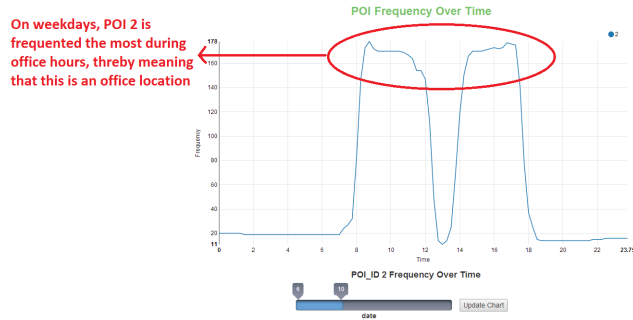Figure 3: POI Distribution over Time Plots

Figure 4: POI Frequency over Time Plots

by employees. These sequences are generated by choosing date, time, POIs and employees. Queries such as *find all the employees from the security department who were present at xyz restaurant on weekdays from noon till 2 pm* can be easily visualized using the interface. Further, the *POI frequency* bar graph and *POI distribution by cars* scatter plot, which are updated for each playback time instance, provides a mapping between cars location and POIs.

2.2 Financial Data Analyzer

A fourth interface (see Figure 2) was designed to analyze financial transaction data that included credit and loyalty card information. The interface provides three different ways to visualize employees' spending patterns: a) *employee vs employee spending comparison* that plots an employee's total spending against all other employees. The comparison can be made either at company level or department level; b) *employee spending distribution* that compares total spending of an employee either across establishments or across days dur-
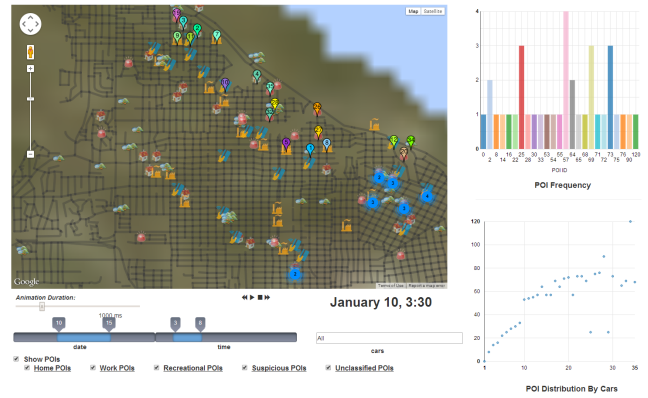

Figure 5: User Location Playback interface

ing the two week window; c) *establishment sales distribution* that compares total sales of an establishment either across employees or across days for the two week window.

3 RESULTS

The POI identification methodology described in Section 2 resulted in 129 POIs that required classification as home, office or recreational. During the classification process, car 9 and car 28 appeared to have broken GPS as there was no trend associated with their location data. These two cars accounted for 64 POI locations visited only by them and no other employee, further strengthening the claim that these POIs are a result of corrupt data and need to be ignored from any further analysis. Of the remaining 65 POIs, 22 were classified as home, 1 as work and 24 as recreational, thereby leaving 18 POIs unclassified that were marked as suspicious. Employees visiting these suspicious POIs were scrutinized further for any questionable behavior or activities. This analysis in conjunction with aberrations identified in employees' spending data provided enough compelling clues necessary for reporting specific individuals as key to the solution to this mini challenge.

REFERENCES

[1] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.