Interactive and Collaborative Visual Analysis on Traffic Sensor Data

Chufan Lai[1]* Qiangqiang Liu[1] Lu Feng[1] Chenglei Yue[1] Xi Chen[1] Yang Hu[1]
Zhanyi Wang[2] Pengju Teng[2] Xiaoru Yuan[1,3] †

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University
2) Qihoo 360 Enterprise Security Group
3) Beijing Engineering Technology Research Center of Virtual Simulation and Visualization, Peking University

Abstract

In VAST Challenge 2017, we propose an interactive and collaborative visual analytic system for the analysis of traffic sensor data. Our system fully incorporates the power of spatial-temporal visualization, sequence mining techniques and collaborative analysis. It allows users to conduct multi-facet and interactive data analysis in a highly efficient way. We discuss technical details in this report, and demonstrate the effectiveness of our system via convincing cases.

1 Introduction

In the Mini-Challenge 1(MC1) of VAST Challenge 2017, the given data is the traffic information recorded by 40 gate sensors in the Boonsong Lekagul Nature Preserve. It is a classic traffic sensor data, where trajectory of each vehicle has been recorded in certain locations. The task is to disclose notable traffic patterns and anomalies taken place in the preserve area. Some unusual behaviours may be of great importance, if they are related to the decline of Rose-crested Blue Pipits.

There are three notable aspects in the data, namely the spatial distribution, the temporal distribution, and the sensor sequences followed by each vehicle. We propose an interactive visual analytic system to fully integrates these three aspects, in order to support an efficient multi-facet analysis of the sensor data. Specifically, frequent pattern mining(FPM) and Dimension Reduction(DR) techniques have been used to help identify patterns and anomalies. Collaborative analysis is also supported to promote communication and collaboration among multiple analysts.

2 Overview

Before further analyses, we first process the data for better pattern mining. Then we introduce 6 views to highlight different aspects of the data, supporting an interactive and collaborative data analysis.

3 Data Preprocessing

In the preprocessing stage, we firstly parse the map image to extract the exact coordinates of all paths and sensors in the park. A color is assigned to each sensor to represent its type.

We reorganize the sensor data by car-ids, resulting in 18,708 vehicles with their sensor sequences in the park. In order to identify similar trajectories, we count the appearances of each vehicle in all 40 sensors. The result is a 40-dimensional data with 18,708 items. Similarity between any two vehicles can be obtained by comparing their sensor vectors in the high-dimensional space.

4 System Components

Our system is a web-based application with 6 views (Figure 1) showing different aspects of the data.

*e-mail: chufan.lai@pku.edu.cn
†e-mail: xiaoru.yuan@pku.edu.cn

4.1 Timeline View

In the timeline view (Figure 1(a)), the histogram shows the amount of vehicles in each day. To enhance small amounts, we allow users to switch the Y axis to the log scale. Users can brush on the time axis, or directly type in to specify a more accurate time range. They can also analyse daily behaviours in the 24-hour timeline.

4.2 Pattern List View

We conduct a frequent pattern mining algorithm to find out subsequences shared by multiple vehicles. Each row in this view (Figure 1(b)) represents a frequent sub-sequence, with its length denoting its 'frequency', i.e. the amount of trajectories (by car-id) that possess such a sub-sequence. All rows are ranked by their frequencies. Users are able to filter the patterns according to their frequencies or sub-sequence lengths.

4.3 Car List View

When the user chooses a frequent pattern, we show the corresponding trajectories in this view (Figure 1(c)). On top of it is the logic sequence of this pattern, while each row in the below is a trajectory possessing this sub-sequence. Colored dots aligned in the temporal order denote different types of sensors pass through by the vehicle. All sequences have been sorted based on their similarity to enhance the visual perception. Users can further choose some trajectories to look for details in the map view.

4.4 Map View

In the map view, we visualize the road network with sensors. When a vehicle is selected, its trajectory is shown on the map in two styles: the route style and the arrow style. The former visualizes the actual route, while the latter clearly displays directions and times of the vehicle passing through this route.

Apart from displaying details, this view also supports users to conduct spatial queries. Users can specify several sensors on the map to find out which vehicles have passed through these sensors. It's a powerful way of query when certain routes are regarded suspicious.

4.5 Projection View

As mentioned above, we have transformed the records into a 40-dimensional data. A simple DR projection (Figure 1(e)) is intuitive enough to show the similarity among vehicles. We choose t-SNE given its power to outline data clusters. Users are able to zoom, brush or highlight a certain type of vehicles in this view.

4.6 Label View

In order to support collaborative analysis, we provide the label view (Figure 1(d)). When users find something valuable, they can create a label with detailed descriptions, and save the findings by assigning this label to the brushed data. Furthermore, they can share their findings with each other by synchronizing the data with a public server. With these functions, users can not only avoid unnecessary repeated searches, but have better communication and cooperation with others.

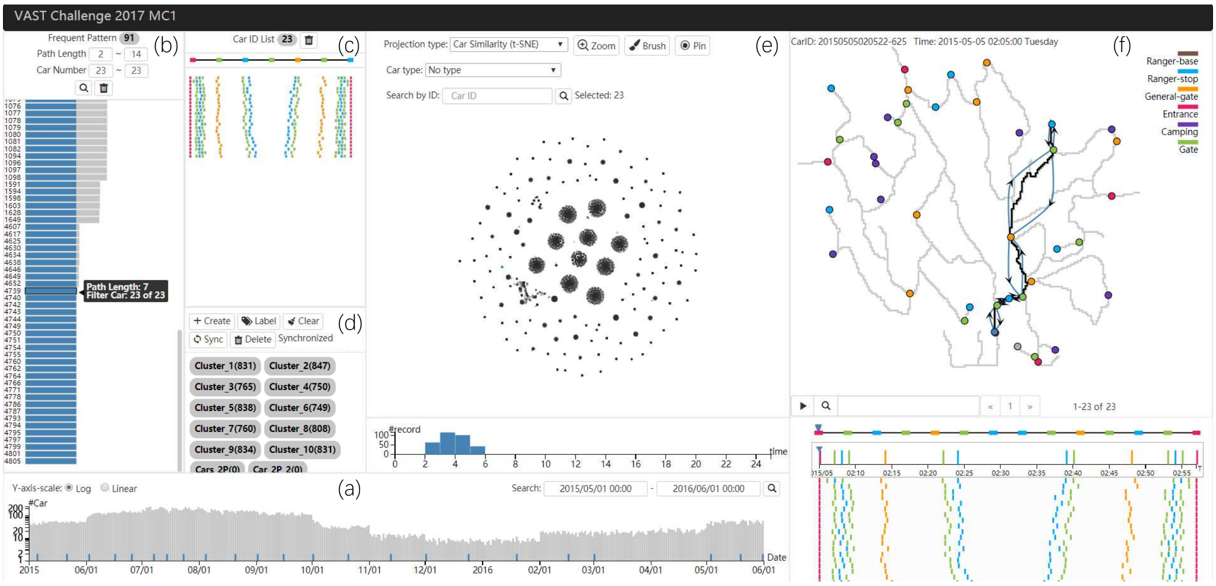

Figure 1: Interface of our system: (a) Timeline View; (b) Pattern List View; (c) Car List View; (d) Label View; (e) Projection View; (f) Map View.

5 Results

The resulting t-SNE is amazingly good, as most vehicles have been tightly included in some cluster (Figure 1(e)). It may due to the fact that the data is an artificial one with many highly similar records. It can also be explained by the inherent spatial limitations. If two vehicles pass through the same set of sensors, they are very likely going the same route. That's because the order between sensors is highly certain given the spatial constraints (i.e. road connectivities) among them.

5.1 Common Patterns

There are 10 major clusters in the projection, denoting 10 major traffic patterns taken place in the preserve area. We've looked into each cluster, and find that these vehicles normally pass through the park from one entrance to another in a very short period of time (less than 1 hour). They are probably the pass-by traffic flow that have no intention of touring or camping in the park.

5.2 Anomalies

We find that most pass-by vehicles have to go through a northern 'bridge' (Figure 2(a)) to travel between the eastern and the western areas. They never take the southern bridge to cut through. But why is that? We soon discover the reason: the southern one is guarded by two Gates, which forbid tourists from passing. Then we filter trajectories that have passed through the Gate bridge. The results are mostly service trucks typed '2P', except for a small group of 4-axle trucks (Figure 2(b)). Who are they?

Combining the other views (Figure 1), we discover that these 23 trucks normally travel during 2:00 to 6:00 in the midnight from a southern entrance to a northern Ranger-stop. Such a behaviour takes place regularly in the whole years. Given all the suspicious features, we identify these vehicles as the 'midnight trucks', label and save them with detailed descriptions. A simple synchronization allows us to share this finding and have further discussions.

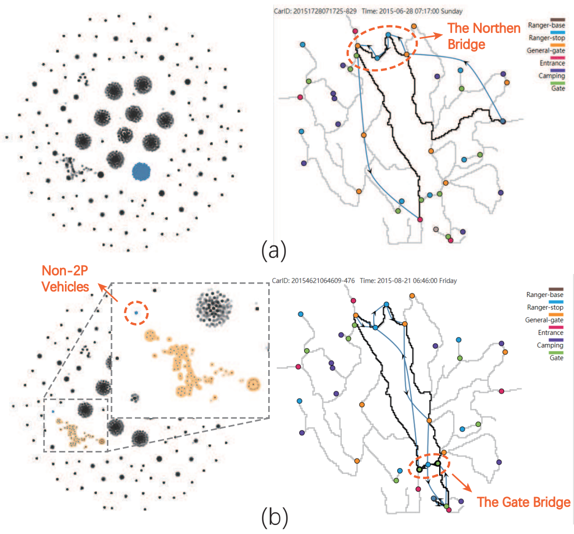

Figure 2: Discovering the unusual midnight-trucks

6 Conclusion

To conclude, we present an interactive and collaborative visual analytic system for multi-facet analysis on traffic sensor data. Specifically, we make full use of FPM and DR techniques to outline potential patterns in the data. With our system, users can efficiently identify common patterns and anomalies in the sensor data, infer the semantics behind the data, and share their findings with each other.

Acknowledgements

The authors wish to thank IEEE VAST Challenge committee and reviewers. This work is supported by NSFC No. 61170204, and partially funded by NSFC Key Project No. 61232012 and the National Program on Key Basic Research Project (973 Program) No. 2015CB352500. This work is also supported by PKU-Qihoo Joint Data Visual Analytics Research Center.