VAST Challenge 2015: Grand Challenge - Team VADER/VIS Award for Outstanding Comprehensive Submission

Michael Steptoe, Robert Krueger, Yifan Zhang, Xing Liang, Rolando Garcia, Sagarika Kadambi, Wei Luo, Thomas Ertl, Ross Maciejewski [*]

ABSTRACT

The VAST 2015 Grand Challenge focused on a fictional park called DinoFun World where visitors are provided with a handheld device for communicating with park patrons and registering their visits to rides. Our goal was to develop a visual analytics system to help explore the spatiotemporal movement and communication data of the park patrons. Specifically, we wanted was to determine which patrons were involved in a theft that occurred at a park pavilion. In order to facilitate this data exploration, we have created a web-deployable system that combines data wrangling, trajectory analysis, network analysis, and interactive visualizations for discovering movement and communication patterns of users and their networks. We earned the Award for Outstanding Comprehensive Submission.

1 INTRODUCTION

Understanding the spatiotemporal movement of individuals and groups is necessary to characterize their behaviors and observe anomalies or unusual patterns. By combining this with communication information we can see the way people act and react based on information received within their network. Several transformations that can be applied to spatiotemporal data to achieve this include trajectory extraction, aggregation, and simplification [1]. In this work, we explore spatiotemporal movement analysis and communications through the use of visual analytics.

2 DATA PRE-PROCESSING

The movement and communication data as provided gave little insight into the parks operations and its visitors. Our approach for processing the movement data was to create a spatiotemporal trajectory structure. The trajectory structure would be defined by attraction check-in information around the park. A large majority of attractions did not have check-in information, but one can infer check-in/out time based on how long a person was idle near an attraction. Inferred check-ins were calculated as follows:

- If a person's location remains within a distance, d, from attraction, a, for more than a temporal threshold, t, then we consider this to be an inferred check-in at a. We used d =5 pixels, $t = 5$ minutes (analysts can vary these parameters).

Our observation is that if users visit attractions in the same order at the same time, then they are likely traveling the park together. We then chose to aggregate the check-ins to 5 minute intervals (user adjustable). A spatiotemporal trajectory also consists of the location that a user last checked into. For example, if the user checks in at the park gate (attraction 84) and then took fifteen minutes to go to the Flying TyrAndrienkos(12) and stayed at that ride for 15 minutes, their trajectory would be: $84-84-84-12-12-12-\dots$

For a full day at the park 8AM-12AM, we have a trajectory of length 192. We combine all three days together for our trajectory

*E-mail: {msteptoe,yzhan212,xliang22,rsgarci1,skadambi,wluo23, rmacieje}@asu.edu; {robert.krueger,thomas.ertl}@vis.uni-stuttgart.de.

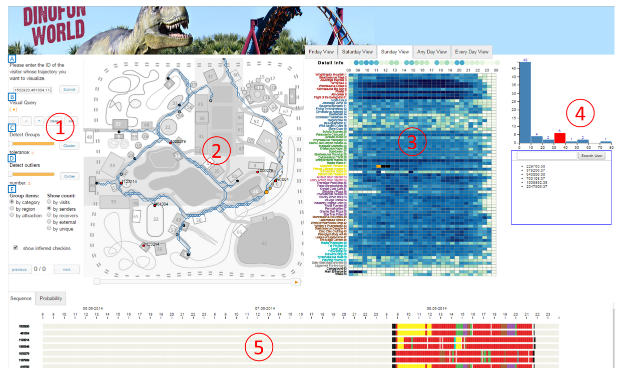

Figure 1: System overview. View (1) shows the analytics interface which allows the user to adjust the system's parameters. View (2) shows a map of DinoFun World and the trajectories of selected visitors. View (3) shows a calendar which represents each attraction in the park and displays the count of visitors in 30 minute intervals. View (4) shows a histogram which displays the number of calls made during a time period. View (5) shows a pixel based representation of a visitor's trajectory in 5 minute intervals.

and a not-in-park value is stored when the user is out of the park. The trajectory structure can be adjusted to represent park regions (five park regions) and attraction categories (thrill ride, etc. - 10 categories total). After creating the structure, patrons with similar trajectories were clustered together using Levenshtein distances (user adjustable). These clusters enabled the exploration and classification of various group types and sizes within the park visitors.

Next we processed the communication data using a spatiotemporal communication structure. The communication structure would be defined by sender/receiver information such as ID and location. From this structure park visitors could then be clustered using their communication networks. Using visitor movement data in combination with their communication data visitors could be aggregated using a blend of network and trajectory clustering. By blending the aggregation techniques the user is able to discover networks of visitor that may or may not spend any time together, but communicate or vice versa. The user can also identify when the visitors within these networks meet, how much time they spend together, and the amount that they communicated.

3 VISUAL ANALYTICS FOR DINOFUN WORLD

For spatiotemporal movement analysis, the trajectory structure previously described is used to visualize, cluster, and map the park visitors' movements through several views. The goal was to enable users to explore individual and group movements interactively to find interesting patterns and behaviors. The system consists of a series of coordinated multiple views and an analytic interface.

3.1 Analytics Interface

The analytics interface allows the user to adjust the system's parameters depending on the objective of their exploration. The interface is made up of five subcontrols.

ID selection: A user can input a list of visitor IDs and view their trajectories on the map view (2). Pressing play will animate the trajectories. The sequence of check-ins for each visitor is displayed in the trajectory view (5). The communications data is also visualized as points on the map that appear at the time of the call.
Visual query: A user can select time and location intervals on the calendar view (2) and create a visual query with logic operators (AND/OR/NOT). This query will return, for example, the IDs of all patron that were at attraction 38 at 4PM and at attraction 45 at 9PM on Friday. This is our primary feature for finding users that were at locations of interest at particular times. IDs and trajectories that return from the query are plotted in the trajectory view (5).
Cluster: All visitor trajectories can be clustered using a Levenshtein distance function and hierarchical clustering. If a tolerance of 0 is selected, the resultant clusters consist of the IDs with identical trajectories (in terms of locations visited at the same time). Reducing the tolerance provides fuzzier clusters (i.e., they have visited mostly the same locations at the same time during their stay). The groups found are plotted in the trajectory view where the trajectory is shown to be the most representative trajectory of the group.
Outliers: The larger the smallest Levenshtein distance is, the more unique a trajectory is. This slider returns the top n-IDs with the largest distance. Results are plotted in the trajectory view.
Calendar Aggregation: This controls how the rows in the calendar view are sorted (by region, attraction or ride type) as well as the data plotted in the cells of the calendar view (data can be the number of visitors at a ride, the number of sent/received/external/unique calls sent from a ride at time t).

3.2 DinoFun World Visualization

The DinoFun World visualization enables the user to examine and interact with the spatiotemporal movement and communication data. Each view manipulates a different aspect of the data and allows for coordinated exploration between each view.
Map View: This view shows the trajectories of selected IDs and also animates their movements over time showing communications during animation and is linked to the trajectory view. The link enables brushing over a section of the trajectory to plot the movement segment on the map. A heat map view coloring each pixel by the number of times a visitor stepped there can also be displayed.
Calendar View: Each row represents an attraction in the park and each cell is colored based on the calendar aggregation control. Data can be viewed for each day, or all three days are aggregated in the any day view. The every day view shows the counts of IDs that were at the same place at the same time every day. Each cell represents 30 minutes of time.
Distribution View: This provides a histogram view of the number of sent/received/external/unique calls made during a time period. The y-axis is the number of IDs and the x-axis is the number of communications made (a histogram of call distribution by ID). Users can click a bin to see all the IDs in a bin, in this way we can find those IDs with unusually large amounts of communications.
Trajectory View: This is a pixel based representation of a visitor's trajectory. Each cell is a 5 minute time interval that is colored based on the location a user is at in the park. By hovering over a cell the attraction or path associated with a patron highlighted on the map.

4 Network Analysis Visualization

Analysis of the spatiotemporal communication data required the examination of visitor communications with respect to location, time, and frequency. This visualization allows the user to see when and where visitors communicate, assemble/disband, and the frequency of user interactions. Input to the view is a list of visitor IDs and the day to examine. The x axis represents the time from 08:00 to 24:00 and the Y axis represents the attractions that the visitors are near. Circles represent the movement of the visitors from one attraction

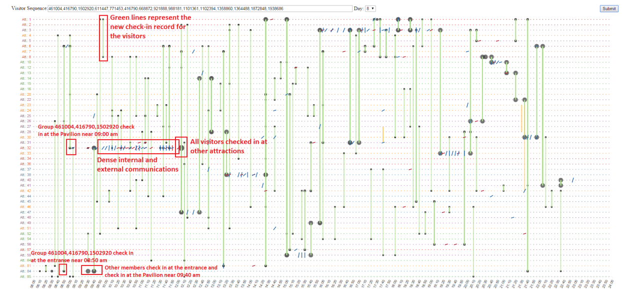

Figure 2: Overview of the network analysis visualization.

to another and the circle size encodes the number of visitors. A circle consists only of visitors that have a link in the communication data. In the lower left, a small black circle can be seen at attraction 84. The small black circle represents multiple people simultaneously checking-in at that time. Later, part of the group at attraction 84 moved to attraction 1. A green line is then drawn to emphasize this movement. By hovering over the line we show if visitors joined or left this group. If the circle changes to red, this means that a person left the group at this attraction. However, if the circle changes to blue a person joined the group. Blue lines represent the internal communications among the visitors and red lines represent their external communications.

5 Contest Results

A famous soccer player, Scott Jones, had his trophies on display for the weekend at the park. During that weekend a crime was committed involving a figure from Scotts past. The goal of the Grand Challenge was to determine which patrons were involved in a theft using a combination of spatiotemporal movement data and spatiotemporal communication data. The challenge asked participants to determine when and where the crime occurred and to identify the possible suspects. Using the calendar view to look at patterns from previous days, we could determine that the crime occurred between 9:45AM and 11:30AM at the Creighton Pavilion. With this information we can identify one group of suspects by creating a visual query that returns all IDs that were in the pavilion during 9:45-11:30AM, for more than 5 minutes (inferred check-ins are 5 minutes). The result of this query was finding several users with inferred check-ins and a single user with a hard check-in.

Next these IDs were used in the Network Analysis Visualization to analyze group movement and communication. From the IDs, we see that they are spatially clustered into three groups. One group waits by an attraction for Scott to pass and ensure he enters the stage at about 9:30. During this time, another group leaves the pavilion and waits at attraction 7 for the group that will enter the pavilion. Between 10AM and 10:55AM a third group is seen at the pavilion while it is supposed to be closed. During this time there is a lot of communication from the pavilion group to the others two groups. Due to this interaction, we hypothesized that the crime takes place between 10AM and 10:55AM. For a demonstration of the system, please see the video at https://youtu.be/LUZr3qEt7Qo.

Acknowledgements

Some of the material presented here was supported by the NSF under Grant No. 1350573 and in part by the U.S. Department of Homeland Securitys VACCINE Center under Award Number 2009-ST-061-CI0001. The authors would like to thank the VAST challenge organizers and reviewers for organizing the contest.

References

[1] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual Analytics of Movement*. Springer, 2013.