Global to Local Pattern of Life Analysis with Tile-Based Visual Analytics

Scott Langevin, David Jonker, Kevin Birk, Chris Bethune, and Nathan Kronenfeld[1]

Uncharted Software, Toronto, Canada

ABSTRACT

In this paper, we present in-progress work on applications of tile-based visual analytics (TBVA) to population pattern of life analysis and geo-temporal event detection. TBVA uses multi-resolution data tiling, analytics and layered high-fidelity visualization to enable interactive multi-scale analysis of billions of records in modern web browsers through techniques made familiar by online map services. Recent experiments show how TBVA can be applied with social media, taxi and weather data to provide insight into day-to-day population movement and activity patterns, and to detect and characterize unusual events in space and time.

Keywords: big data, visual analytics, pattern of life analysis, geo-temporal event detection.

Index Terms: H.5.2 [Computer Graphics]: Graphical User Interfaces (GUI).

1. INTRODUCTION

The term "big data" has become a ubiquitous reference to data sets which are large and complex enough to cause traditional analytic approaches to fail. Discovering and understanding geo-temporal patterns and anomalies at scale is one such problem. Computational techniques for detecting statistically significant data events exist, but characterization and assessment of relevance remains an issue. Insight into complex geo-temporal patterns is also limited.

In this proof of concept application for geo-temporal analysis, TBVA [1] is used with computational analytics to correlate geo-located New York City area Twitter posts, taxi pick-ups and drop-offs and local weather records to provide insights into pattern of life at global and local scales. The application is currently undergoing instrumented experiments where users must complete tasks such as identifying the time and route of a marathon or protest by the observable effect it has on patterns of movement and social media activity.

2. TILE-BASED VISUAL ANALYTICS

TBVA uses Aperture Tiles [2] cluster computing technology to bin raw data and compute derivative analytics in a power-of-two pyramid of data tiles, similar to how a map service such as Open Street Maps (OSM) pre-computes and serves image tiles [3]. TBVA dynamically serves data tiles to both server and client-side layer renderers, providing capabilities for end-user filtering and correlation.

3. PICKUP AND DROPOFF PATTERNS

Applying TBVA to 187M taxi pickups and drop-offs reveals distinct patterns in taxi trips in and out of the city, including local hot spots, GPS signal diffusion between skyscrapers, and direction of travel indicated by side of street (Figure 1).

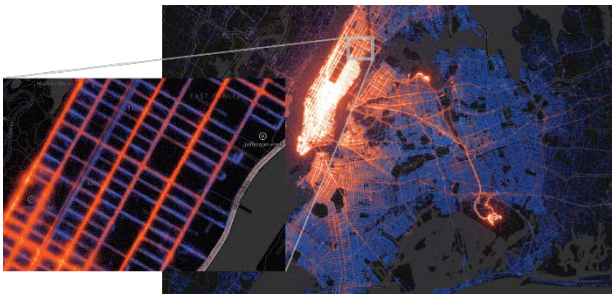

Figure 1: Patterns of taxi pickups (red) and drop-offs (blue).

Computing and displaying variance from normal trip patterns reveals anomalous behaviors, which can be correlated with social media activity to discover events such as a street festival (Figure 2). Both top hashtag and tweet-based topic modeling are used to convey population discourse and for event characterization.

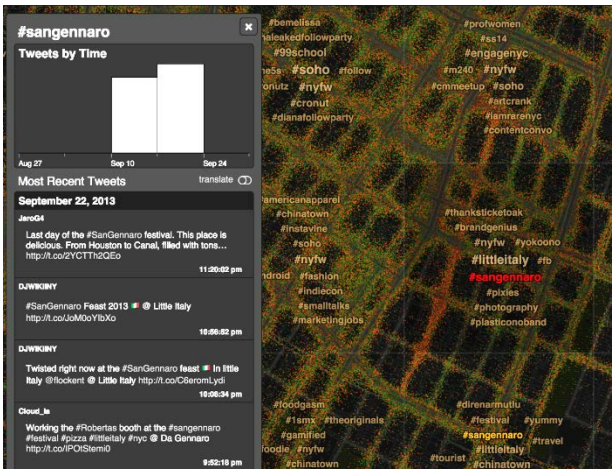

Figure 2: Correlating a change in traffic observed in red with social media reveals the San Gennaro street festival.

Traffic modeling is also used to compute and display a cross-plot of expected trip length over time. Correlating spikes in trip time with an interactive histogram of weather event summaries points to potential causes, such as snowstorms.

4. SOCIAL MEDIA PATTERNS

Segmenting Twitter data to identify a subpopulation of New York City tourists and inferring trips from geo-located posts reveals popular movement patterns and destinations when links are tiled between post locations (Figure 3). Long extrapolations however can often dominate, interfering with visual detection of more localized paths and events. Marking the results of computational geo-temporal event detection in context complements visual methods of discovery, making it easier to spot interesting phenomena (Figure 4).

[1] slangevin@uncharted.software, djonker@uncharted.software, kbirk@uncharted.software, cbethune@uncharted.software, nkronenfeld@uncharted.software

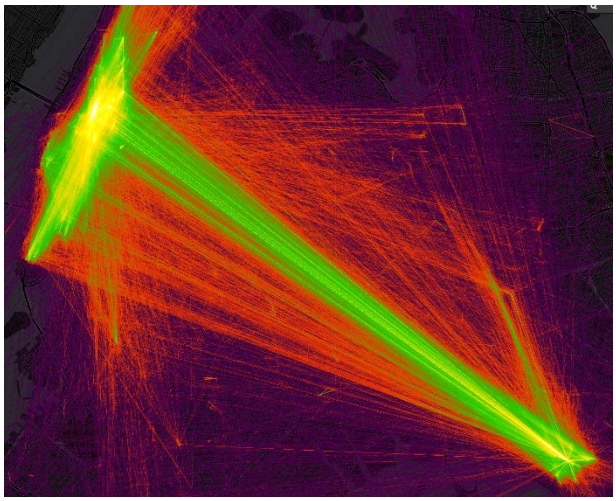

Figure 3: Inferring tourist trips from social media posts reveals patterns of travel and popular destinations in New York City. The color scale increases luminance through hue to show increasing density [4].

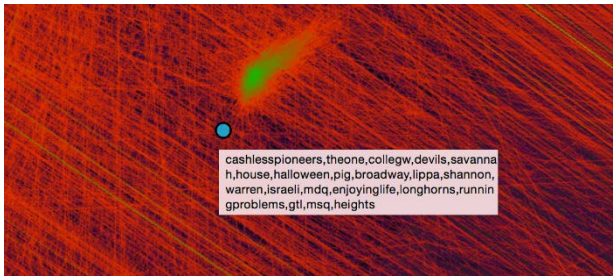

Figure 4: Computational event detection and marking in context aids discovery while fully expressing the event visually.

5. EVENT QUERY

Capabilities for event query are also provided in the application. Twitter keyword search coupled with automated query term expansion and executed in ElasticSearch produces social media hits which are plotted on top of existing layers, enabling an analyst to rapidly find events of interest. A query by example capability is also provided where the user is able to draw a boundary around an event or physical entity such as a hospital or park, and submit the associated social media data as training data for finding similar phenomena. Query results are then clustered for display to facilitate browsing of results (Figure 5).

Figure 5: Social media Query by Example techniques are used to find similar events or places such as marathons, or hospitals.

6. USER TESTING

User experiments with more than 20 people are ongoing in preparation for deployment to the analytic community, with promising results thus far. Instrumentation captures analyst workflow as they complete challenge tasks using the tool, such as finding marathon routes, protests, and points of ingress for tourists. With no more than fifteen minutes of introduction to the tool, a large majority of users with an average level of industry experience or higher complete the tasks successfully with a high degree of confidence in a short amount of time. Agility of the tool is also being evaluated formally through the late introduction of new data sets for analysis. Final evaluation results will be included in poster publication.

7. CONCLUSIONS AND FURTHER WORK

TBVA has shown to be a highly promising technique for analyzing pattern of life at macro and micro population scales across billions of data records. Highly scalable, layered, high-fidelity visual analytics provide a powerful method of finding and understanding geo-temporal patterns and events. Search in context has shown to be an invaluable capability for analysts to rapidly find answers to relevant questions.

Further work remains in refining techniques for visual summary of social media topics to optimize cognition of patterns and anomalies. Higher level tailored view constructs must also be introduced to simplify control of layers. Experiments with non-linear opacity along links may also prove fruitful in reducing visual interference of long spans exhibited by inferred tourist trips, while reinforcing uncertainty about route.

10. ACKNOWLEDGEMENTS

This work was supported by the Defense Advanced Research Projects Agency (DARPA) under Contract Number FA8750-12-C-0317. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official Department of Defense position, policy, or decision.

REFERENCES

[1] Cheng, D., et al. Tile Based Visual Analytics for Twitter Big Data Exploratory Analysis. IEEE Big Data Conference, 2013.

[2] Aperture Tiles, 2014. URL: http://aperturetiles.com/.

[3] Potmesil, M., Maps Alive: Viewing Geospatial Information on the WWW, Computer Networks and ISDN Systems 29.8, 1997.

[4] Ware, C., Information Visualization: Perception for Design. Elsevier, 2012.