Visual analytics support for collecting and correlating evidence for intelligence analysis

Siming Chen
Peking University

Chenglong Wang
Peking University

Zipeng Liu
Peking University

Zhenhuang Wang
Peking University

Zuchao Wang
Peking University

Zhengjie Miao
Peking University

Xiaoru Yuan*
Peking University

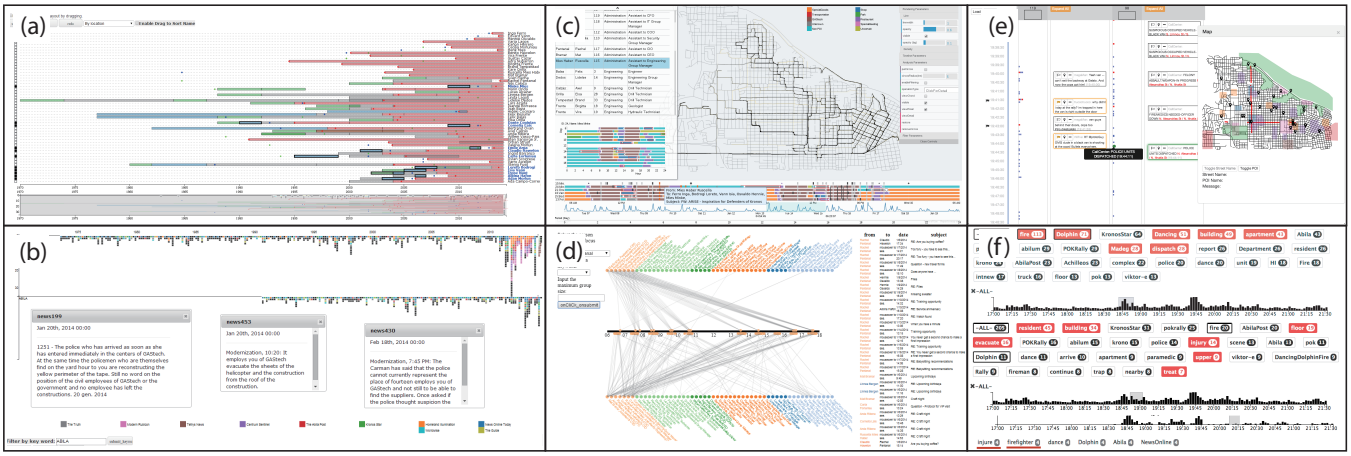

Figure 1: System interface, including the (a) News Timeline, (b) CV Timeline, (c) Spatial-temporal View, (d) Email View, (e) Streaming Tweet View, (f) Keywords View.

Abstract

Understanding multiple types of data source and correlating them make intelligence analysis a challenging task. Visual analytics techniques help analysts better comprehend the data and find correlations. We proposed a visual analytics system mainly designed for collecting and correlating evidence in intelligence analysis. Based on our system, we provided a analytical solution for IEEE VAST Challenge 2014 Grand Challenge. With the system, analysts collect information from spatial, temporal and network data. Multiple correlating methods are provided to help analysts generate a wide range of hypotheses and test their reliability.

1 Introduction

Intelligence analysis is an important application area for visual analytics [2]. Currently, the analysts in intelligence analysis are facing larger amounts of heterogeneous dataset. Gorg et.al proposed a series of work on intelligence analysis with the aid of visual analytics [1]. Pirolli and Card [3] summarised the cognitive tasks in intelligence analysis. Specifically, our work aims to solve the critical two parts - collecting information and correlating pieces of evidence. With these functions, our visual analytics system could help analysts generate a wide range of hypotheses and verify them. Our system has the following advantages in helping the intelligence analysis. 1) Collecting and integrating heterogeneous data. 2) Interactive outlier detection. 3) Events correlating and hypothesis generation.

*e-mail: {csm, wangchenglong, zipeng.liu, wangzhenhuang, zuchao.wang, miaozhengjie, xiaoru.yuan}@pku.edu.cn

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

2 Tasks and Data Description

IEEE VAST Challenge 2014 provided a scenario for intelligence analysis. The scenario is in a fictitious country Kronos. In Kronos, a big company called GAStech was experiencing a kidnap. An organization known as the Protectors of Kronos (POK) is suspected. Participants of the challenge are required to figure out what happened, who was doing the bad things, what were the motivations and high-level network behind the scene with visual analytics.

The data for analysis includes the news data, worksheets and documents of GAStech employees' CVs, email headers of the employees, GPS logs and transaction logs of employees, streaming tweets and catalogs of police and fire department. Thus as a analysts, we need to find special events and generate hypotheses based on the text, network, spatial temporal and streaming social media information.

3 Visual Analytics Process

Our visual analytics system consists of six views (Figure 1). In the exploration phase, we also took advantage of the JigSaw [1], Python, QGIS, etc. The pipeline includes four iterative stages, including general pattern detection, outlier detection, event correlation and hypotheses generation (Figure 2).

3.1 General Pattern Analysis and Event Detection

News and CV data These types of data involve time series information and text descriptions. With NLP techniques, we extracted people, organization, time as entities from each text files. In the CV timeline (Figure 1a), each employee's experience was stacked along the y axis. The color encodes different education, military and working experience. Thus general pattern of each one's experience is shown. Having detected special experience, users can sort and select a group of people for detail comparison. In the news timeline, each news article is shown as a rectangle, the layout of

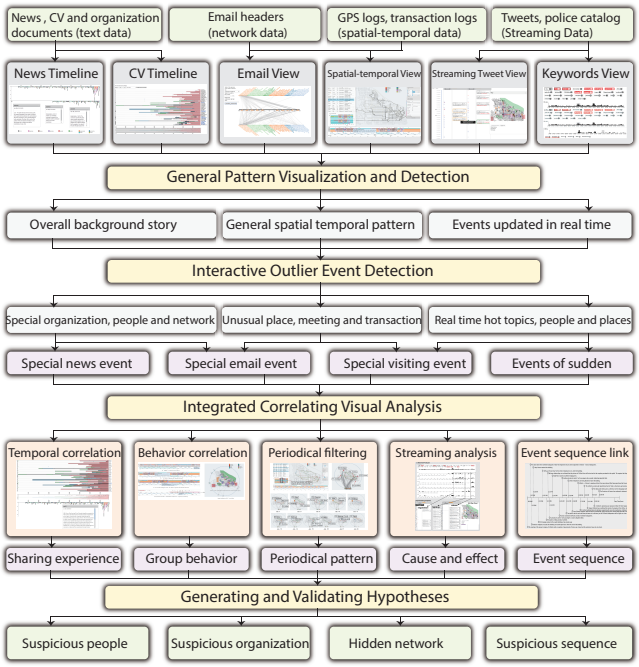

Figure 2: System pipeline.

which is based on the time. To understand the background information, users can select each article to view the text and search the keywords. Special news events reflected by the search results would generate a new timeline. Thus we can extract the important events along the time, with the main characters and organizations.

GPS logs, transaction logs and email data Our system integrates these behavior data into the spatial temporal view and provides interactive filters (Figure 1c). The spatial view shows the spatial distribution of GAStech employees' trajectories. We can select one or multiple persons to compare their routes in spatial view and event timeline. Transaction log and email headers are visualized as circle and line along the GPS event timeline (Figure 1c bottom). Daily pattern of email, trajectories and transactions can be analyzed together through the interface, which reduce the analyzing cost of shifting the focus. To further investigate the email behavior, users can query the topics and people in the email view (Figure 1d). The top layer indicates the senders while the bottom layer indicates the receivers. The axis in the middle indicates the email timeline. Thus based on the understanding of the behavior pattern, we could target to the following types of abnormal events: visiting a strange location, visiting a location at strange time, gathering of employees, large transactions and emails with special topics or people.

Streaming tweets, police and fire department logs In the streaming data view, we can see each tweet falling down along the vertical timeline (Figure 1e). Details can be extracted and GPS related information would automatically tagged on the map. In the real streaming situation, our interface allow multiple analysts analyzing the tweets collaboratively. Users are able to create customized data filters based on source, topics, keywords to get a sub-stream. Control center can assign each filtered streams to each person for dividing the tasks. Keywords view (Figure 1f) works as an supplemental view to understand the dynamic keywords change. With time filter function, users can easily know the hot keywords and characters that people are talking about in different period of time as general patterns. Once mastering the trends, they can drill down to the tweets content for some bursting topics or emergency events.

3.2 Event correlation and hypotheses generation

After collecting multiple pieces of events as evidence, our system provides several mechanism for analysts to correlate them.

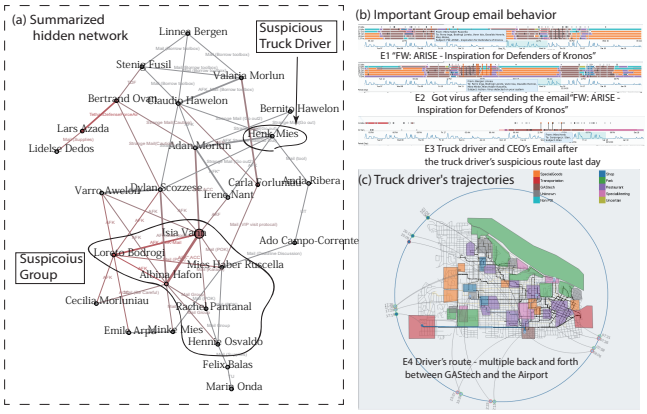

Figure 3: Illustration of Event Correlation. (a) Extracted hidden network. (b) Email events. (c) Spatial temporal event.

Temporal correlation and event sequence linkage In the CV view, analysts can select a period of time with specific experience. People who shares the experience could be queried. Through the exploration, we can explore the hidden network based on sharing experience, eg. classmates, former working partners. The hidden network contributes to part of the overall network (Figure 3a). Analysts can verify them based on other events happening in the same group of people. For example, we could detect the group of people sending suspicious emails titled "POK". Afterwards the GAStech network is been affected by virus (Figure 3b - E1, E2).

Behavior correlation In the spatial-temporal view, once we identified the suspicious places, we can filter correlated people who also went there. We can also examined the other behaviors of the these people, including the email and transactions. For example, we found one truck driver was communicating with CEO through multiple emails (Figure 3b - E3). One day before this event, the truck driver had suspicious trajectories indicated by the spatial temporal view (Figure 3c). We could correlate these events together and investigate the relationship between the truck driver and CEO.

Periodical filtering Based on periodical filtering, analysts can get email sending, trajectories and transaction pattern. We could also detect a group of people who shares the same suspicious pattern.

Streaming analysis In streaming analysis of the Tweets and catalog data, we could correlate the events based on temporal order. We could also refer to the spatial view to investigate whether some new events happen in the suspicious places we've already detected.

4 Future Work

In this work, we developed a visual analytics system to support the intelligence analysis. Currently, due to the space limitation, users need to change contexts in each view. In the future, we would deploy the system into the tiled display wall and make it fully linked.

Acknowledgements

The authors wish to thank IEEE VAST Challenge committee and reviewers. This work is supported by NSFC No.61170204

References

[1] C. Gorg, Y. ah Kang, Z. Liu, and J. Stasko. Visual analytics support for intelligence analysis. *Computer*, 46(7):30–38, July 2013.

[2] T. J.J. and C. K.A. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005.

[3] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. of International Conference on Intelligence Analysis*, 2005.