VAST 2014 Mini-Challenge 1: MEAT – Multiview Event Analysis Tool of Diverse Data Sources

Zhuang Cai[*] Mengyao Chen[*] Hanqing Zhao[*] Ying Zhao[†] Fangfang Zhou[†] Kang Zhang[‡,*]

[*]Tianjin University School, [†]Central South University, [‡]The University of Texas at Dallas

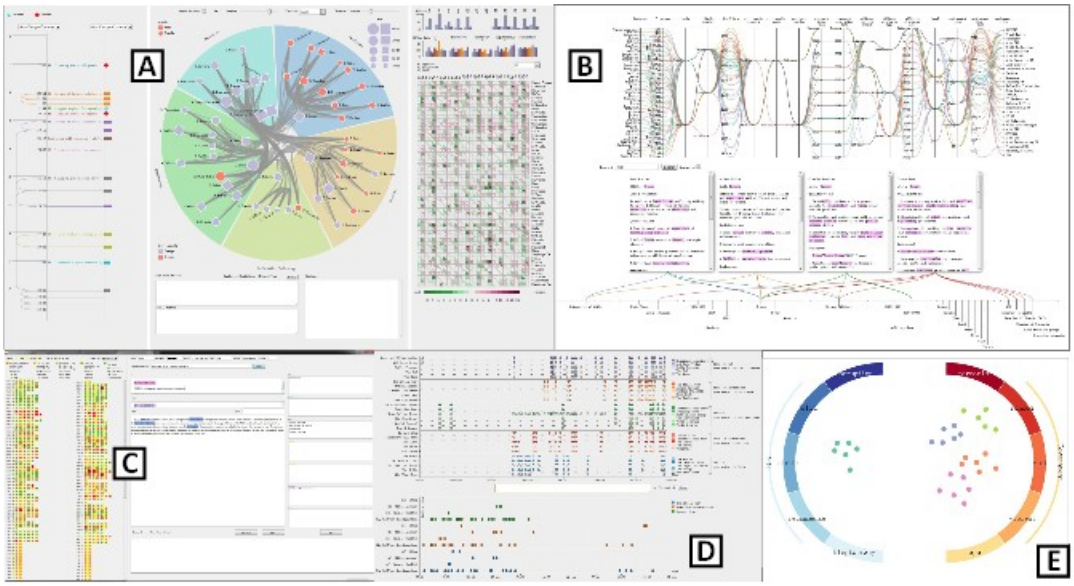

Figure 1: Overview of MEAT visual intelligence analysis system.

ABSTRACT

Dealing with diverse data sources has becoming a new challenge in visual analytics. MEAT intelligence analysis system is designed to meet the challenge of VAST 2014 Mini-Challenge 1 [1]. Using a coordinated multiview with a clear interaction pipeline, MEAT supports complicated analyzing tasks, including intelligence collection and reasoning. Using MEAT, we have successfully mined a great amount of information from the available data, resolved the problems one after another and finally delivered a convincing outcome.

Keywords: visual intelligence analytics, diverse data sources, coordinated multi-view visualization, visual reasoning.

Index Terms: H5.2 [Information Interfaces and Presentation]: User Interfaces—GUI; I.2.7 [Artificial Intelligence]: Natural Language Processing—Text Analysis; K.4.3 [Automation]: Anomaly Detection—Visual Hints

1 INTRODUCTION

Data collection, analysis and reasoning are essential for knowledge discovery. To comprehensively grasp the backgrounds and motivations of targeted criminals, law-enforcements need to handle multisource data potentially full of conflicts. The VAST 2014 Mini Challenge 1 includes a dataset of emails, resumes and news etc. The task is to find evidences about kidnapping in an island country, Kronos, through visual analysis. Our work focuses on helping law-enforcement assess the situation to figure out where the missing employees were by mining the crime network over time and analyzing their communication details.

Our tool, MEAT, is designed to analyze multi-source heterogeneous data visually using coordinated multiple views and an interactive pipeline. It integrates sporadic events to provide an explicit view from implicit connections.

2 DATA AND PROCESSING

In context of the VAST 2014 MC1 [1], 7 kinds of data were provided in a variety of formats, e.g. .csv, .doc, .txt and .xlsx at from across the past 20 years, 1170 email headers of two weeks within GAStech and a spreadsheet of employee records with partial resumes. For data pre-processing, we used the method of td-idf vector for articles clustering, such as when the news were reported, who was the reporter and from which media the news were released. Entity extraction was also used for speed-reading and information analysis. Furthermore, the credibility of information sources and media outlets plays important roles. Thus, we picked some biased keywords like kleptocracy and ransom to obtain the word frequency in the reports of each media, which helps in tendency analysis in Radvis (Fig 1.E) [3].

3 MEAT OVERVIEW

As shown in Figure 1, MEAT consists of four different views in 3 parts: PizzaVis, P2CVIS, ACKVis and ATMVis (Fig 1.A-D), to be introduced in this section.

* email: caizhuang16@gmail.com

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

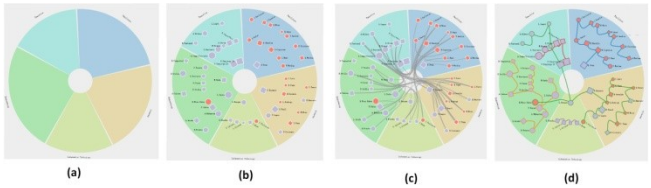

Figure 2: Four layers of PizzaVis.

3.1 PizzaVis

As the data available are merely the email headers without contents, our work focuses on the GAStech structure and employees' relationships. PizzaVis provides three views for visualizing the email relationships, including their time distribution and subjects. The 'Pizza' in the center of Fig 1.A has four layers as shown in Fig 2. The bottom layer shows the 5 departments of GAStech in Fig 2.a. In Fig 2.b, the second layer shows a hierarchical structure of employees. The third layer uses hierarchical edge bundling to visualize the email relationships [2]. As shown in Fig 2.d, the top layer is optional, showing color-coded clustered information. The left sub-view represents email communications between any two selected employees. By threading emails of the same topic using arcs, investigators can easily see the time and topic streams, and thus infer the relationship between the two individuals. The matrix view on the right shows email statistics, revealing the active periods of each employee.

3.2 P2CVis

P2CVis (personal profile and connection visualization) is designed to show employee structures with discrete attributes shown in Fig 1.B. The parallel coordinates above represent structured features of employees. Analysts can query and select any individual by interactively brushing, or view his/her related properties. In the middle of P2CVis, filtered resumes are displayed horizontally. By identifying and highlighting keywords, one can reveal employees' unstructured attributes. From these resumes, analysts can get an overview of staff's experience, expertise and special skills. Furthermore, they extract anyone of interest with specific features by filtering. For example, if analysts want to check out who controls the fire alarm, they can filter the word 'alarm'. The bottom of P2CVis shows potential connections between employees and POK, such as common military services.

3.3 ACKVis and ATMVis

In Fig 2.C and 2.D, Article clustering and keyword visualization (ACKVis) and article time and media visualization (ATMVis) are built for article analysis. ACKVis focuses on analyzing raw news data, while ATMVis focuses on inference.

ACKVis shows small squares representing articles on the left, color-coded for time as well as media dimensions. Article squares are also temporally clustered. We highlight key words using an entity extraction method. ACKVis also support keyword editing and abstraction, allowing quick summarization of events in the news. Having distilled the key events, we proceed to reason about them using ATMVis.

Based on scatterplot, ATMVis is aimed at simplifying the analyst's cognitive load by globally displaying news with their release media and release time. Through keyword filtering, the analyst can rapidly locate relevant events and construct the logical sequence. Depicting media bias, ATMVis can further infer incident causes and consequences.

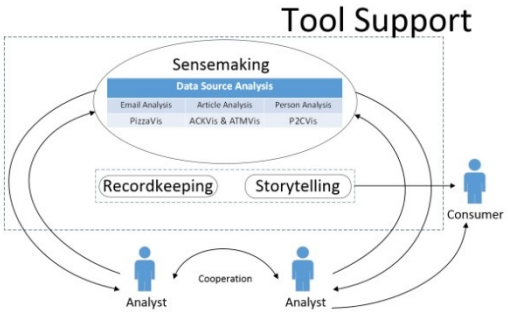

Figure 3: Framework of visual intelligence analysis of MEAT.

4 Analysis Process

Using MEAT, we detect and analyze the POK structure and its status during the kidnapping and important events happened on that very day. We also reason about the kidnappers' motivations. Figure 3 shows a framework of intelligence analysis. A detection example about a suspicious mail and collaborative analysis are discussed in this section.

4.1 Collaborative Analysis

One crucial aim of MEAT is providing a shared tool support for information flow to support an intelligent investigative process. The main views of MEAT are developed using D3. The Web-based tool is convenient for analysts to share information anywhere at any time. Snipping function is also provided by integrating investigations into a database to support future reasoning and storytelling. MEAT has a historical record searching function for case sharing so that every analyst can learn from his/her partners.

4.2 Example – a Mysterious Email

As an example, we start with Isia Vann. Double clicking "Isia Vann", all related employees are highlighted. By filtering mass e-mails, we find that Isia Vann and assistant Rachel Pantanal were closely related. On the left view, e-mails only show their special relationship. Next, by clicking on R.Mies.Haber, we find a suspicious e-mail with five recipients, referring to POK and ARISE. We conduct subject filtering to find who else received this e-mail. Pie charts show dangerous employees, such as Bodrogi, Osvaldo, Ferro etc., and close ties are found between them. Meanwhile, by clustering private relationships, we find that Linnea Bergen, Isia Vann and Osvaldo were in the same cluster. Clicking on this three individuals, the e-mail on the 14th with the same subject "Action: Virus detected~" shows up. Further, group members had a heated discussion on the two e-mails. Linnea Bergen was the manager of IT Department that had unusual number of e-mails on that day. Moving the mouse to the matrix view, we can see emails of every IT employee. Clicking on employees, e-mails between them are displayed clearly. We speculate that this e-mail was forwarded by Haber to a suspicious group of five people and triggered the investigation by the IT Department. It may also have motivated Bergen to send an alert email.

5 Conclusion

With multisource data, intelligent analysis is a challenge. MEAT uses innovative multi-view and clear interaction pipeline to help analysts work together. It supports deep mining of valuable information from given data, providing a solid basis for further detection. MEAT needs more powerful functions,

such as summarizing and storytelling. These will be our future work.

References

[1] IEEE, http://www.vacommunity.org/VAST+Challenge+2014

[2] Holten D. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data[J]. Visualization and Computer Graphics, IEEE Transactions on, 2006, 12(5): 741-748.

[3] Sharko J, Grinstein G, Marx K A. Vectorized radviz and its application to multiple cluster datasets[J]. Visualization and Computer Graphics, IEEE Transactions on, 2008, 14(6): 1444-1427.