SMART: Social Media Analytics and Reporting Toolkit

Jiawei Zhang\textsuperscript{*} \hspace{1cm} Junghoon Chae\textsuperscript{‡} \hspace{1cm} Chittayong Surakitbanharn\textsuperscript{‡} \hspace{1cm} David S. Ebert\textsuperscript{§}

\textsuperscript{*}Purdue University \hspace{1cm} \textsuperscript{‡}Oak Ridge National Laboratory \hspace{1cm} \textsuperscript{§}Purdue University

\textbf{ABSTRACT}

We present the Social Media Analytics and Reporting Toolkit (SMART), a web-based visual analytics system that enables the end users to effectively identify actionable information and gain situational awareness from social media channels. The development of SMART has been guided by an iterative design process and close collaborations between visualization researchers and emergency responders. SMART provides real-time social media analysis through topic extraction, cluster examination, anomaly detection and message categorization. These components are integrated into a visual and interactive interface that allows the users to navigate, supervise and customize the exploration of real-time streams. The system has been deployed and used by several law enforcement agencies in several national and regional events for the purpose of real-time monitoring and emergency management.

\textbf{Keywords:} Visual Analytics, Social Media, Situational Awareness

\textbf{Index Terms:} H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical User Interfaces (GUI)

\section{INTRODUCTION}

Social media have become an important source of information and situational awareness in crisis and emergency management [10, 13, 15]. With the massive number of messages generated and diffused through online social media platforms, locating meaningful and actionable information in a timely manner is crucial, yet non-trivial for decision makers [1]. Automatic data mining algorithms in this field are commonly designed to solve fine-grained tasks such as topic classification or anomaly detection [7, 11, 12]. However, they lack the ability to support exploratory data analysis and incorporate domain knowledge into the analysis loop [16].

In this paper, we present a system called the Social Media Analytics and Reporting Toolkit (SMART), which is an integration of our previously published research [3–6, 14]. SMART follows general visual analytics principles [16] and aims to provide a scalable and interactive analysis environment by combining automatic data mining modules and web-based visual dialogs. As a human-in-the-loop framework, SMART consists of layered visual presentations for conveying heterogeneous information dimensions including space, time and semantic content in a holistic visual space. SMART provides highly customizable analysis components such as user-defined topic classifiers that allow for task-tailored exploration and navigation within the dynamic and complex analysis space. SMART leverages the client-server architecture, with the compute-intensive tasks deployed in the back-end server for interactive performance.

SMART has been deployed to multiple public safety and law enforcement agencies including police departments, U.S. Coast Guard sectors and fusion centers for real-time monitoring and emergency management. The effectiveness of SMART has been demonstrated through use during multiple real-world events including 1) large-scale planned events such as conventions and festivals; 2) recurring events such as football games; and 3) abnormal events such as crimes and natural emergencies. SMART has also been incorporated into several college courses for introducing visual analytics and its application in social media analysis into curriculum in emergency management and disaster analysis. Currently, the system is not open sourced but we are glad to provide access for users who are interested in exploring the system.

\section{DESIGN AND DEVELOPMENT ITERATIONS}

The project kicked off when our law enforcement partners expressed their need to utilize social media for emergency management in Indiana Public Consortium meeting in 2011. At the beginning stage, SMART was adapted based on Scatterblogs, a desktop software developed by our partner university [3, 14]. However, the early feedback from our partners indicated that the tedious installation and configuration of the desktop version made its use problematic. Therefore, we refactored the framework into a web-based platform for better accessibility over the past four years.

A user-centered design process has been adopted in the development of SMART. Two major groups of experts participated in the process, including the technique development team (us) and the public safety law enforcement partners (intelligence analysts). Our partners had varied domain backgrounds ranging from police departments to Department of Homeland Security fusion centers to the United States Coast Guard (Figure 3). During the major development stage, we conducted a series of iterative sessions with our partners. These sessions usually occurred before, during and after a specific event in which our partners wanted to use SMART to monitor social media in real time. The sessions generally followed the following procedure: One to two weeks before the event, we provided in person or online training and a FAQ session to familiarize the end users with the SMART system and the newly developed features. This would especially benefit the end users who used the tool for the first time or who have not used the tool for a relatively long time, typically a few months. During the event, the end users used the system and recorded any issues or suggestions related to SMART. Meanwhile, we provided remote support to resolve any technical issues. Once the event ended, we conducted a review session (hot wash) with the end users to evaluate the utility of the system during the event and identify strengths and limitations. The review session was mainly guided by interviewing them based on the following major questions:

- Which features were effective or mostly used? Did the feature really help identify actionable information, or just because it was visually appealing?
- Which features were not useful? Was the interface too complicated and difficult to use? Did the visualization provide misleading information? Is there any improvement to resolve this issue?
- Which new features would potentially be useful? Could they be implemented by extending the current components? Or do they need to be built from scratch?

\textsuperscript{*}e-mail: zhan1486@purdue.edu
\textsuperscript{‡}e-mail: chaej@ornl.gov
\textsuperscript{§}e-mail: ebertd@purdue.edu
As Figure 1 shows, the primary workspace of SMART employs a geographic view since space has been regarded as the most critical dimension for situational awareness according to our partners. For the information related to the geographic dimension, we use a layered visualization approach that stacks different visual presentations on top of the geographic background. For information not directly related to a spatial location or that is infeasible to stack on a map, we have created separate visual dialogs that are integrated with the map view through the brushing and linking design pattern. The end users can easily toggle on or off the display of map layers and visual dialog boxes based on their needs. The current system supports geo-located Twitter and Instagram data that are publicly available.

3 System Overview

As Figure 1 shows, the primary workspace of SMART employs a geographic view since space has been regarded as the most critical dimension for situational awareness according to our partners. For the information related to the geographic dimension, we use a layered visualization approach that stacks different visual presentations on top of the geographic background. For information not directly related to a spatial location or that is infeasible to stack on a map, we have created separate visual dialogs that are integrated with the map view through the brushing and linking design pattern. The end users can easily toggle on or off the display of map layers and visual dialog boxes based on their needs. The current system supports geo-located Twitter and Instagram data that are publicly available.

3.1 Layered Visual Representations

SMART consists of a set of layers on top of the map view to visually convey spatial, temporal and semantic information. The point layer visualizes each message as a single dot and applies a hot-dot animation to the newly posted messages in order to draw the user’s attention. Moreover, the newly posted messages have a high opacity and old messages have a high transparency. Messages that are outside the pre-defined time window (e.g., 2 hours) fade out as they are not of interest to analysts during real-time monitoring. A heat map visualization is supported in the system that visualizes the spatial hotspots. Extensions of SMART has focused on analyzing individuals’ trajectories patterns and spatial distribution across multiple scales. More details can be found in [4, 18].

To visualize textual features on the map, SMART has applied two major techniques: a Tag Map (Figure 2(left)) and a Contentlens (Figure 2(right)) [3]. The Tag Map generates an occlusion-free spatial text layout to indicate the prominence of keywords in different regions. It is a static visualization that presents a high-level semantic abstraction and does not require user interaction. In contrast, the Contentlens allows the users to interactively move a circular lens on the map, with prominent keywords extracted from the points within the lens and visualized based on a word cloud layout near the lens.

3.2 Interactive Message Categorization

SMART provides an interactive and customizable component for message categorization as shown in Figure 1(d) [3]. In this classifier view, each circular node represents a specific topic and associates with a set of relevant keywords. The keywords can be generated either based on automatic classification algorithms or defined by the end users interactively in the system. The node shown on the map indicates the volume of microblogs relevant to the topic and updates in real-time, which intuitively reflects the prominence of different topics on social media. When the user clicks the node, the system filters the messages relevant to this topic.

The classifier view supports a set of logical operations including union (OR), intersection (ADD) and inverse (NOT). A union operation combines the messages from two nodes into a new node. An intersection operation groups the messages that are in both nodes into a new node. The nodes generated by logical operations are rendered in blue and yellow, respectively, in order to visually distinguish from the original nodes. These logical operations enable the users to extend the current classifiers by using intuitive interaction means, thus reducing the tedious operations required to define new nodes. An inverse node (rendered in grey) is defined during the node creation stage and groups messages that do not contain the keywords associated with this node.

3.3 Abnormal Topic Detection and Automatic Alerts

SMART incorporates LDA topic modeling [2] and visualizes them in a word cloud style (Figure 1(c)) [3]. When the analyst clicks a single keyword, the system filters and shows the relevant messages. SMART applies Seasonal-Trend Decomposition (STL) [8] to quantify the degree of anomaly of the topics [5]. The analyst can sort and observe the top-ranked abnormal topics and investigate the details in other coordinated dialogs.

SMART adopts an interactive and exploratory analysis paradigm, thus requiring active user involvement during the analysis process. As a supplementary component, SMART provides automatic and customizable alert functionality, which aims to provide real-time situational awareness when an active user involvement is infeasible or expensive. Based on the user-defined geofence and keywords, the module automatically alerts users when the volume of the relevant messages hits a threshold within a specific time window. The users can also configure the module to send email summaries including relevant information on a regular basis.
4 USE CASES AND FEEDBACK

As Figure 3 shows, our end users have used SMART for monitoring multiple national and regional events. In this section, we present several use cases of SMART. We then provide domain expert feedback and general discussion.

4.1 Using SMART for Planned Events

We use the RNC event as an example to describe the general usage (Figure 1). One week prior to the event, the analysts provided us with an event-specific profile including a targeted region for monitoring and several groups of keywords in which the analysts were potentially interested. We then added the region in the predefined list in SMART and associated the keywords with their user accounts. The keyword groups mainly included a security-relevant group (e.g., bomb, explosion, gun), a protest-relevant group (e.g., protest, activist, boycott) and a event-specific group (e.g., convention, GOP, Cleveland). In the mean time, the email services were enabled based on these keywords. While the keywords were usually defined based on the end users’ domain knowledge and previous events, they could adjust and improve them according to the actual topics occurring in the social media streams as monitoring occurred.

During monitoring, the analyst started the analysis by examining the trending topics in the topic view and reading the volume of different classifiers she defined early on. She identified massive job-related posts and then enabled the NOT filter to reduce the noisy information. By examining the heat map, she found a prominent hotspot containing RNC-related topics formed and grew around Cleveland, where the convention was hosted. In contrast, nearby regions had more topics related to traffic and entertainment.

Hence, she drew a geofence around Cleveland and zoomed in to gain improved situational awareness related to the convention.

She then turned on the security-related classifier, with the topic view updated to show relevant topics such as police, arrest and helicopter. This reflected the fact that a lot of protesters were at the convention and there were several conflicts with the police officers. She then observed several hotspots near Public Square and Quicken Loans Arena, indicating the major regions of the on-going protests. In order to gain a holistic picture and avoid missing important information, she combined the security-related and protest-related classifiers and filtered based on the new node. At the same time, she drew a geofence to only focus on those locations with large crowds. With more and more messages posted in the selected region, she identified new topics such as pencarry and blacklivesmatter. As these new topics were not included in the initial keyword list, she added them in the protest-related node to capture more relevant posts that were not expected prior to the event. This helped her understand the evolving situation more comprehensively.

4.2 Capturing Unexpected Issues Using SMART

SMART has successfully provided situational awareness in previous events. During the riverboat festival in Louisville (October 2014), the email service captured a suspicious tweet that said "EVERYONE SKIP SCHOOL TOMORROW RT @xxx: IM LITERALLY GOING TO KILL SOMEONE" (the user id is anonymized). The intelligence analysts were able to identify the user and his/her location and dispatch officers to follow up on the issue. During Purdue football games, the police officers using SMART identified several users that had violent behaviors and positioned officers to prevent inci-
The trajectory analysis module in SMART was used to monitor the human movement and traffic condition during the Battle of Bristol football game in Tennessee. The visual results clearly revealed the traffic flow patterns including the major directions of the movement and the transportation type (walking or driving a car). According to the analyst, the visual outcome corroborated the reality on the ground where certain areas were experiencing traffic with a significantly higher volume. Major keywords were highlighted on the map along with the traffic flows to provide contextual information.

4.3 Feedback and Discussion

In general, our end users had very positive feedback for SMART. They all acknowledged the importance of utilizing social media for emergency management and public safety. They especially emphasized the need for an immersive exploratory environment that allows them to customize and supervise the monitoring and analysis in an interactive manner. They welcomed the fact that SMART adopted various visual presentations to show multiple information dimensions contained in social media data. At a high-level perspective, the analysts found useful that the system allowed them to observe an overview of the current situation, customize filtering to narrow down to an analysis subspace and examine specific users or tweets on demand. The analysts reported liking the interface design that used the geographic map as the major workspace and stacked other supporting dialogs on top of the map, as the geographic dimension was the most important information dimension for their situational awareness. One end user also reported that SMART allowed him to understand the mood of the crowd during the event, which he believed could be an important first indication of potential issues.

The end users also had positive feedback on the individual components in SMART. They commented that the content lens was both visually appealing and useful. The tool enabled them to quickly observe the trending topics in different regions at different spatial scales, compare the prominent topics at those regions and potentially identify the abnormal events. They liked the classifier view that allowed them to customize keywords based on their monitoring in real time and quickly review and understand whether the changes were reasonable. They agreed that interactively manipulating the classifier nodes and performing logical operations on them were straight-forward and efficient. They had positive feedback for the NOT classifier that effectively removed a massive amount of daily chatters and highlighted more meaningful actionable information.

SMART mainly consists of web-based interfaces that were refactored based on a desktop version. The web platform provides rich visualization libraries and a flexible environment that accelerates the development process. However, the limited computational resources on the browser side (e.g., Google Chrome has a 4GB memory limit on 64-bit machines) require a back-end support to handle compute-intensive modules to ensure the interactivity of the front-end interface. For high-performance rendering on the browser side using WebGL, one has to consider that WebGL provides a subset of OpenGL features and could potentially limit the use of advanced rendering functionality. Furthermore, various types of browsers/versions the end users may use and different security settings their machines have should also be taken into consideration to ensure the successful delivery of the system.

As a potential limitation, publicly available and geo-located social media data can potentially be sparse during small-scale events. While we try to infer the location of the non-geo data from the message content [9], as future work, we plan on exploring and integrating other social media data sources that may be available to use. This would help the analysts observer the users’ activities from multiple social media channels, thus enriching the analysis space available to them. Moreover, real-time social media streams are highly dynamic and evolving, which may introduce abrupt changes in the visual space and increase cognitive overload to the analysts. Extended upon our previous research on spatial context preservation [17], we plan on investigating potential solutions for maintaining the context of consecutive visual states in the temporal dimension, thus helping analysts to effectively understand the evolving situations.

5 Conclusion and Future Work

We have presented SMART, a visual analytics environment that utilizes real-time social media data for emergency management and public safety. SMART is a collaborative outcome between visualization researchers and public safety law enforcement analysts, and provides interactive and customizable visual components for effective social media analysis. We have demonstrated the efficacy of SMART through several real-world usage scenarios.

As future work, we aim to combine other publicly available social media data sources into the system in order to enrich the analysis space and potentially provide cross validation in the analysis process. As real-time streams are highly dynamic, we plan on exploring effective approaches for reducing abrupt changes of dynamic visualizations and maintaining the context across consecutive visual states.

Acknowledgments

This work was funded by the U.S. Department of Homeland Security VACCINE Center under Award Number 2009-ST-061-CI0003.
References


