NStreamAware: Real-Time Visual Analytics for Data Streams (VAST Challenge 2014 MC3)

Fabian Fischer*
University of Konstanz

Florian Stoffel†
University of Konstanz

Figure 1: The image at the top represents a part of the data stream using our sliding slices visualization, which summarizes the stream using sliding windows to provide a summary timeline. The colored histogram at the bottom highlights major events based on extracted keywords and insights of interesting events which were identified by the analyst in real-time.

Abstract

To solve the VAST Challenge 2014 MC3 we use *NStreamAware*, which is our real-time visual analytics system to analyze data streams. We make use of various modern technologies like Apache Spark and others to provide high scalability and incorporate new technologies and show their use within visual analytics applications. Furthermore, we developed a web application, called *NVisAware*, to analyze and visualize data streams to help the analyst to focus on the most important time segments. We extracted so-called *sliding slices*, which are aggregated summaries calculated on a sliding window and represent them in a small-multiple like visualization containing various small visualizations (e.g., word clouds) to present an overview of the current time segment. We show how these techniques can be used to successfully solve the given tasks.

1 Introduction

The fictional scenario of VAST Challenge 2014[1] was the so-called *Kronos Incident* in which several employees of a company named *GAStech*, located at the island of *Kronos* went missing. Because of an ongoing conflict between an organization known as the *Protectors of Kronos (POK)*, they are suspected in the disappearance.

Within that challenge, the main focus of MC3 was to analzye a real-time data stream based on (1) microblog records that have been identified by automated filters as being potentially relevant to the ongoing incident and (2) text transcripts of emergency dispatches by the local police and fire departments. The overall task of the challenge is to analyze the data stream in real-time and identify interesting events to help the law enforcement from Kronos to assess the situation and figure out where the missing employees are and how to get them home again.

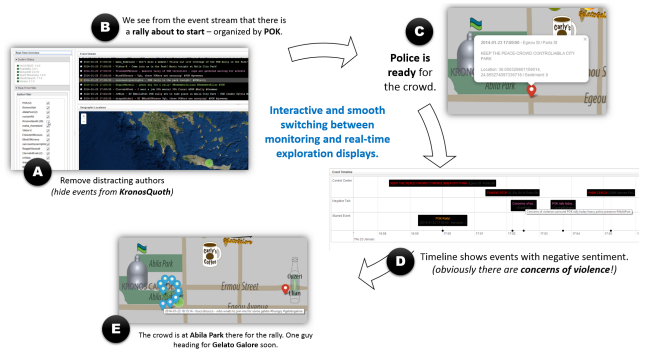

Figure 2: Workflow of interactive switching between various displays and central management of interesting events, which got starred by the analyst, in a common timeline.

*e-mail: Fabian.Fischer@uni-konstanz.de
†e-mail: Florian.Stoffel@uni-konstanz.de

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 System Architecture

To process the data stream, we made use of various modern technologies to provide a scalable infrastructure for our modular visual analytics system, which is called *NStreamAware*. Our architecture consists of our *REST Service*, *Spark Service* and *NVisAware*. To provide proven and scalable data processing, we make use of Apache Spark [1], RabbitMQ[2], ElasticSearch[3], and MongoDB[4].

The *REST Service* connects to the data streams and preprocesses the data and gathers various additional information for the incoming events. The sentiment value for each message is calculated, topic classification is applied and street names or intersections are mapped to the most likely GPS coordinates. The service does also provide a REST interface to retrieve historical data or manage insights. All events are stored to a distributed ElasticSearch cluster and are forwarded to our message broker RabbitMQ and to our Spark Service to generate real-time summaries on sliding windows and stores them to a MongoDB database. We call these summaries, which are generated in a regular interval, *sliding slices*. Those slices and also a selection of raw messages are forwarded to our web application *NVisAware*, so that they can be visualized in the graphical user interface to the analyst using various interactive real-time displays similar to the approach published in [2]. As seen in Figure 2 (A), the events are shown as textual list, but also in a geographic map based on the gathered coordinates. Within the textual list, which is colored according the sentiment value, it is possible to click a star icon to store the message to the insights management timeline for future reference. Various real-time filters can restrict the shown messages to different authors. The geographic map supports zooming and panning and plotted messages are clustered to avoid overplotting. The clusters can be clicked to expand them to follow all messages originating from a particular location.

The messages are also pushed to our *Spark Service*, which runs on top of the Apache Spark Streaming [1] platform for analytics. Spark Streaming is optimized for the use in large distributed cluster environments to provide scalability even in big data scenarios.

To reduce the analysts' cognitive load, we implemented a visualization for our sliding slices as seen in Figure 1. A single slice contains summarized information about most frequent words, main topics, talking authors, new authors, which have never been seen before and a similarity score to previous slices, which is mapped to the background color. The word cloud helps to get an idea of the major topics discussed in the current time slice.

3 Some Results

During the real-time analysis, we identified various interesting events, which are briefly discussed in the following. All insights have been generated using *NVisAware*. In Figure 1 the timeline at the bottom shows the evolution of various main events over time.

3.1 POK Rally with different Speakers

The real-time display and also the sliding slices make it clear, that a rally organized by POK is about to start and will be present the first two hours of the data stream, which can be backed up by various starred messages.

```
17:00:00 - POK rally to start...
17:02:31 - Concerns of violence.
17:27:00 - End of Lucio's speech.
19:02:00 - Marek thanks speakers...
19:07:00 - POK rally ends in a concert...
```

[2]http://www.rabbitmq.com/
[3]http://www.elasticsearch.org/
[4]http://www.mongodb.com/

The workflow of switching between different displays during real-time analysis and putting messages to the insights timeline can be seen in Figure 2.

3.2 Dolphin Apartment Fire

The sliding slices as seen in Figure 3 together with the geographic map help to easily identify a major fire in the Dancing Dolphin apartment complex around 18:40. After the fire flares up again at 21:00, there is even an explosion at around 21:30.

Figure 3: Sliding slices before and during the start of the fire.

3.3 Black Van, Hostages, and Shooting

Figure 4 highlights the start of a situation where a black van, with hostages on board, is involved in various accidents and can eventually be stopped by the police near Gelato Galore. A police officer got shot and the SWAT team arrives. At one point the black van drivers surrender and the hostages are free again.

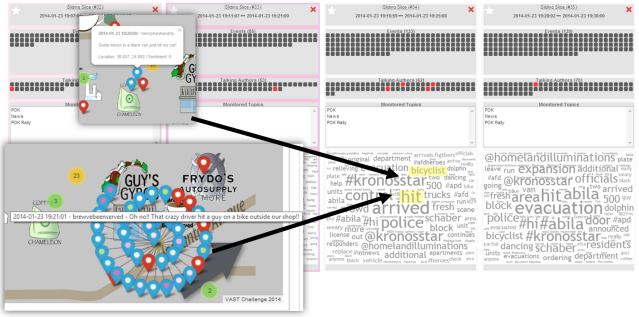

Figure 4: Identification of a suspicious black van hitting a car and a biker. The black van can be stopped near Gelato Galore.

4 Conclusions

We showed in the VAST Challenge 2014, how *NStreamAware* can successfully be applied to a real-time data stream and how *NVisAware* and our approach of using visual sliding slices can be used to identify suspicious events and achieve good situational awareness for previously unknown data streams.

References

[1] Apache. Spark Streaming. https://spark.apache.org/streaming/.

[2] F. Fischer, F. Mansmann, and D. A. Keim. Real-Time Visual Analytics for Event Data Streams. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, SAC '12, pages 801–806, New York, NY, USA, 2012. ACM.