Spatio-temporal Visual Analysis for Event-Specific Tweets

Mashaal Musleh[§]
KACST GIS Technology Innovation Center, Umm Al-Qura University, Makkah, KSA
mmusleh@gistic.org

1. INTRODUCTION

Twitter is one of the most popular social networks where people use to tweet about their opinions, feelings, desires, ...etc. One of the most important and consistent behaviors of Twitter users is posting a plethora of tweets about *events* of different types, e.g., Oscars celebration, soccer games, and natural disasters. For such kind of event-specific tweets, geo-tagged tweets grab the biggest attention because all events by nature have a spatial extent. For example, while Boston Marathon explosions were going on, in April 2013, users rush to Twitter seeking tweets from the marathon location [2]. Thus, within a large project called Taghreed for comprehensive real-time and offline analysis and visualization for Twitter data (see www.gistic.org/taghreed), we are working on a module that aims to semi-automate the task of visually analyzing event-specific tweets, for arbitrary events, over the spatial and temporal dimensions. In this poster, we present our on-going work on this module and discuss three of its use cases. We also discuss our future plans to extend the module for larger data sizes and make it interactive while the events are going-on.

2. PROBLEM STATEMENT

Input: A set of tweets that are related to a specific event.
Output: A web-based application that facilitates visual spatio-temporal analysis of the input tweets.
Constraints: The application should enable:

- Interactivity, so that any operation's response time should not exceed fractions of a second.
- Spatial analysis at any granularity level within the event spatial extent.
- Temporal analysis at two levels: single point of time and time range within the event period.

With our problem statement constraints, our ultimate vision is to provide an easy tool to create an *efficient* and detailed *visual* analysis tool for a large number of tweets, overcoming the existing limitations of visualization tools. For example, in Oscars 2014 night, the famous Bradley Cooper selfie is retweeted 3+ Million times making a new retweet record. The previous record, of 750K+, was achieved by Barack Obama's "four more years" tweet which is also related to the event of USA presidential elections.

3. RELATED WORK

The majority of work on events in Twitter community lies in the category of event detection through conducting spatio-temporal analysis on incoming tweets, few examples are [1, 4, 5]. The only work that addresses automating event analysis is [3]. However, this work mainly targets the problem of automating the analysis of users' responses to identify the segments and topics of the event that got praise or criticism. Both types of work are inherently different from our goal to automate producing visual analysis user interface. In fact, the output of both can be an input to our module to automatically generate the visual analysis interface.

4. SOLUTION OVERVIEW

As we highlighted before, our module aims to *semi-automate* the visual analysis task for a specific event. That points to existence of non-automated steps in the whole process. In this section, we give a brief overview of the whole process. First, we describe the non-automated data collection step for a specific event. Then, we describe how the current module takes this data and produces the corresponding web application per the problem statement in Section 2. Finally, we discuss the future plans for this module to handle larger data sizes and being interactive while the event is going-on.

Before running our module, we manually collect geotagged tweets, through Twitter public streaming APIs, that are related to the target event. In this context, events data are classified into three categories based on event type: (1) Event data that is defined only with certain hashtags, e.g., Oscars, with no certain spatial extent as many people tweet about it globally, (2) Event data that is defined only with certain spatial extent, e.g., Hajj which is a Mulsim religious trip that happens annually only in certain regions in city of Makkah, Saudi Arabia, and (3) Event data that is defined with both certain hashtags and spatial extent, e.g., local sports games. Around the clock, Taghreed, the project that our module is part of, is drinking from Twitter streaming APIs and archiving the tweets in a secondary storage. We extract the event data from either the archive or the streaming APIs directly if we know ahead about the event.

§The author is mentored by Amr Magdy and Mohamed F. Mokbel who are affiliated to both University of Minnesota and KACST GIS Technology Innovation Center at the time this poster is published.

SIGMOD'14, June 22–27, 2014, Snowbird, UT, USA.
ACM 978-1-4503-2376-5/14/06.
http://dx.doi.org/10.1145/2588555.2612666.

After collecting the event-specific tweets, they are stored in files and fed as an input to our module. The module is producing a simple web application that currently consists of a single HTML file and a single JavaScript file. The web application interface has two mandatory components and two optional components. The two mandatory components are: (1) A geographical map, currently based on Google Maps APIs, that display the spatial distribution of the tweets, and (2) A timeline that allows to move from a time slice to another. The geographical map has two modes of operation: it either presents the individual tweets along with their text content or a heatmap that shows the relative density of the tweets over the space. The timeline has three modes of operation: selecting tweets at either one point of time, a temporal range, or play the tweets changes over time to show the changes in a chronological order. The two optional components are: (1) A check-box set that optionally exclude tweets of certain areas, e.g., cities, and (2) A statistics panel that is used only in what called *two-side events*, where the event involves tweets about two competitor entities, e.g., sports games or elections, to display statistics about tweets of each side. The module takes a configuration file that determines the input tweet files, the included optional components along with its configuration, and the temporal granularity of the timeline. The spatial and temporal extents are automatically detected from the tweets and Google Maps by default allow the zooming functionality to explore the data on different levels of spatial granularity. Section 5 is discussing actual use cases with different examples for the produced web interface.

The future plans for this module lies in two folds: (1) Handling larger sizes of data, and (2) Producing interactive web applications that changes while the event is going-on. Currently, we are using Twitter public streaming APIs to collect data which gives a small percentage of the whole firehose. Thus, the largest number of event-specific tweets we collected so far is 40K tweets. Currently, we handle tweets in main memory for interactive response with the users. However, if we want to scale for the whole firehose size, we can get an extreme example like Oscar 2014 celebration where more than 3 Million tweets are related. For such number, additional challenges are imposed in both using the main memory of web browsers and for rendering the data on the geographical map layers. Our plan is to extend this to use a sort of secondary storage that renders data on geotagged images in the back end. Then, we send those images to the front end to be aligned with the geographical map tiles. This should keep the interactivity requirement satisfied.

As we are using live streaming APIs for data collection, a possible and interesting extension for our module is to produce a visual analysis interface for future events where the data collection is running live while the event is going-on and fed to the web application periodically, e.g., in hourly basis, so that the analysis changes online. Once the event comes to an end, the analysis interface would remain stable as if it is produced with the current version of the module. This may be interesting for some events like USA presidential elections where the whole nation, and probably the whole World, is sitting in front of screens following and tweeting the results.

5. SUMMARY OF MAJOR RESULTS

We applied our module to three major events: (1) Oscars 2014 celebration, (2) Saudi final cup soccer game, and (3) Hajj, which is the largest gathering of Muslim people worldwide annually in city of Makkah, Saudi Arabia. Figures 1 and 2 show two examples of the produced interface. We can see the difference between a two-side event like the soccer game and a normal event like Oscars 2014. In a two-side event, the statistics panel is used to show ratio of tweets of each side over time. Also, the soccer game is a local event that make sense to use check-boxes to include/exclude tweets from certain cities in Saudi Arabia. The three produced interfaces are available online on www.gistic.org/oscars2014, www.gistic.org/hilal_vs_nasr, and www.gistic.org/hajj-tweets.

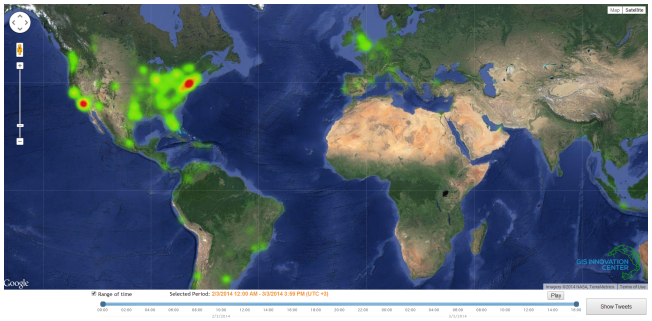

Figure 1: Oscars 2014 tweets

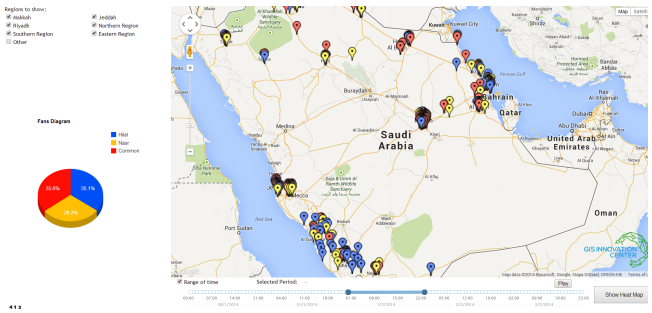

Figure 2: Saudi final cup soccer game tweets

The provided spatio-temporal visual analysis is helpful in showing the users behavior during the event. For our use cases, in Hajj tweets, our interface tells something about the flow of pedestrians during Hajj days in Makkah regions and people opinions about the provided services, through showing individual tweets, which is important for the administrative authorities. In the local soccer game, one can visually measure the popularity of each team by city and over time. This is useful for sports analysis broadcasts.

6. REFERENCES

[1] H. Abdelhaq, C. Sengstock, and M. Gertz. EvenTweet: Online Localized Event Detection from Twitter. In *VLDB*, 2013.

[2] http://www.latimes.com/business/technology/la-fi-tn-after-boston-explosions-people-rush-to-twitter-for-breaking-news-20130415,0,3729783.story,2013.

[3] Y. Hu, F. Wang, and S. Kambhampati. Listening to the Crowd: Automated Analysis of Events via Aggregated Twitter Sentiment. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2640–2646, 2013.

[4] R. Li, K. H. Lei, R. Khadiwala, and K. C.-C. Chang. TEDAS: A Twitter-based Event Detection and Analysis System. In *ICDE*, 2012.

[5] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *WWW*, 2010.