Storms and Floods in Social Media: How the @INSIGHT project follows extreme weather phenomena

Sylvia Steenhoek1, Gennady Andrienko2, Natalia Andrienko2, Ioannis Katakis3, Antonia Saravanou3, George Valkanas3, Dimitrios Gunopulos3, George Fuchs4, and Hendrik Stange4

1 Federal Office of Civil Protection and Disaster Assistance (BBK), Germany
2 Fraunhofer IAIS, Germany and City University London
3 National and Kapodistrian University of Athens
4 Fraunhofer IAIS, Germany

Social media are at the spotlight of civil protection organizations and academia since they comprise a great source of information enabling the development of applications with social impact. Recently, Twitter data are exploited in a plethora of software tools aiming at better emergency response. One of those efforts is the INSIGHT project, a European Collaborative project whose goal is the utilization of urban data for effective and efficient disaster management. We demonstrate the potential of Twitter analysis in two scenarios related to severe weather situations.

Following Floods. Early identification of weather incidents as well as monitoring affected areas are important milestones in providing aid and allocating resources. Interesting results were obtained from a case study on the floods that occurred in the United Kingdom during January 2014. We collected Tweets from the area by considering their GPS footprint and their content. The geographical segmentation of the whole area into regions - defined as polygons - was done by applying a technique known as ‘clustering’ which groups together areas with similar characteristics. This analysis enabled the visualization and ranking of the most affected areas (see Figure 1) based on Twitter posts. Human annotators confirmed the efficacy of the approach by contrasting its output with what really happened, using information from the U.K. Met Office.

An interesting observation was that areas that were affected similarly by the extreme weather condition (when it started, when it reached its peak and when it finally settled down) presented similar data trends of flood-related tweets (see

*An edited version of this article is published in the magazine ‘Crisis Prevention’ (Beta Verlag und Marketinggesellschaft mbH), Issue 4 (2015), in the German language (Stürme und Hochwasser in Social Media: Wie das INSIGHT Projekt #extreme @Wetterphänomene beobachtet)
Figure 1: Most affected areas using several metrics (total number of tweets, total number of flood-related tweets, flood related tweets ratio (SNR)). On the right, a map of the actual rainfalls. Correlation between affected areas as identified by INSIGHT (especially in the third case - SNR) and the ground truth data indicate that social media can help towards the analysis of extreme weather phenomena.

Figure 2). For this analysis, we extracted spatio-temporal features for each area from the Twitter data. Subsequently, we grouped areas together based on that information, using, once again, clustering. Each generated group contained areas where the (flooding) event manifested similarly in time and space.

**Tracking Storms in Germany.** In the same spirit, we used a set of georeferenced Twitter messages posted on the territory of Germany on October 27-29, 2013 in relation to the storm Christian, a.k.a. St. Jude storm. The goal here is to evaluate the potential of tracking such a large scale phenomenon by analyzing a social media stream.

From all tweets posted in Germany during that period, potentially relevant ones were selected by finding occurrences of keywords indicating extreme weather conditions. Only 491 messages have been found; hence, the event stream is sparse in both space and time. Generally, the Twitter users in Germany post much fewer messages than users in the UK. For this case study, ground truth was available from the mass media and the experiences of the researchers. We utilized a visualization tool built by our research group that analyzes these tweets and groups them together in space and time. This means that messages that were posted in nearby areas and close in time will be placed in the same ‘cluster’.

We checked if the detected clusters were located in the affected areas and if the existence times of the clusters correspond to the times when the storm
reached these areas. We run the tool and see that nothing significant happened in Germany on 27/10/2013, Sunday, except for a small cluster of tweets in Berlin. We inspect the texts of the tweets, which can be interactively accessed, and see that they either occasionally contain some of the keywords or refer to events occurring in the UK and the Netherlands.

In Figure 3 below, there are 6 screenshots of the situation display corresponding to different 3-hour time intervals on 28/10/2013, when the storm reached Germany. The red circles show the current positions of the cluster centers and the total numbers of events in the clusters. In the lower left corner of each map we have put an inset with a fragment of the map legend showing the number of tweets represented by the largest circle size. We observe that this number increases over time.

As in the previous day, a cluster emerged in Berlin, but the messages did not really refer to the situation in the city. A similar cluster appeared in Munich. The cluster in Frankfurt contains mostly messages concerning flight delays due to the storm or turbulences during landing at the airport. The remaining clusters and their evolution correspond to the path of the storm over Germany: it first hit the west of Germany and then moved northeast to Schleswig Holstein, Lower Saxony, and Hamburg, where the largest cluster was detected. On the
next day, the storm was over. A single cluster emerged in Hamburg with tweets referring to the storm of the day before.

This study demonstrates the ability of our tools to detect significant clusters in a sparse event stream. It also underlines the necessity to exercise caution when trying to use georeferenced social media for locating extreme events or affected areas. First, the positions of the posts may be misleading since the users can react to events happening elsewhere. Second, clusters of posts are more likely to occur in highly populated places. The sizes of the clusters may correlate with the population and may not be indicative of the severity of the local conditions.

Our experiences in the INSIGHT project revealed the great potential of Social Media in disaster management. However, this is only one source of information. Modern cities and countries are a vast with the data (sensors, vehicles, mobile phones, etc). The future of emergency monitoring lies in the analysis and combination of heterogeneous data streams.

Relevant Publications