Combining Supervised and Unsupervised Learning to Detect and Semantically Aggregate Crisis-Related Twitter Content

Jan Bongard[1] (jan.bongard@dlr.de), Jens Kersten[1]
[1]German Aerospace Center (DLR), Jena, Germany

Twitter for Disaster Management:

Social media have a considerable potential to complement usual information structures in crisis situations [1]. Every device that is connected to the social network can serve as a multimedia sensor. Thus, text messages, images, videos or geo-locations can be used to improve the situational awareness of political officers or relief forces [Fig.1].

Twitter data is suitable because of its worldwide active user community and its free Stream-API provides live access to 1% of all public posted tweets. But currently, the integration of social media streams into the decision-making process of disaster management is accompanied by several hurdles. These arise mainly due to the information overload and the low reliability of individual tweets.

Sufficient information on natural conditions is often available from satellite or drone images. Therefore, information on health or social conflicts is of strong interest. Since time is an important factor in disaster management, it is essential to filter the small amount of important content from the voluminous stream in near real time. In addition, the data must be visualized in a clear and interactive way. However, crises can change quickly, so data filtering and visualization must be able to adapt to the momentary demands of the user.

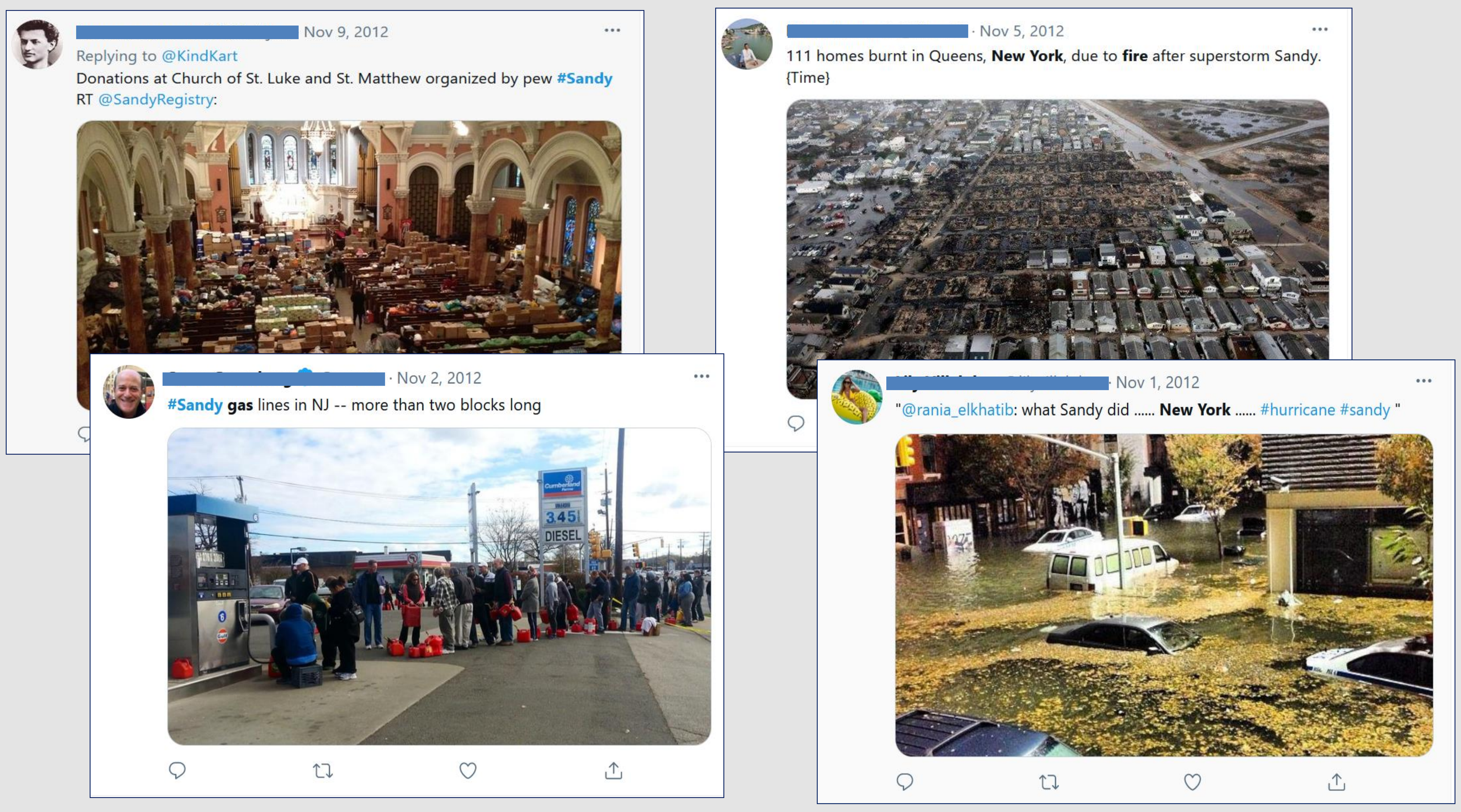

Fig. 1: Selection of tweets from hurricane Sandy, New York, November 2012

Objectives:

State of the art filtering techniques are often based on supervised machine learning techniques [1,2]. Although they process individual tweets very efficiently, adequate training of the model its adaptation to unseen event types is very costly. Unique subevents or unforeseen developments of a crisis cannot be trained in advance, so there is a risk of ignoring crisis-relevant tweets. Unsupervised learning considers tweets from the past and evolves with the data stream. Furthermore, it can be customized to the specific needs of different disasters. Drawbacks are the higher computational effort and the need of user interaction for result interpretation. The intention for combining both methods is to:

1) Enhance the generalization capability of pre-trained models
2) To be able to handle massive amounts of stream data
3) Reduce information overload by identifying potentially crisis-related content
4) To obtain a semantically aggregated data representation that allows further automated manual and visual analyses

Method:

The basic idea is to divide the stream into intervals [Fig.2]. The tweets of each interval are preprocessed and converted into a numerical vector using Google's Universal Sentence Encoder [3]. Then, Chinese Restaurant Process Clustering is performed using cosine similarity as distance measure. Instead on filtering single tweets, the decision about relevant or not relevant can now be made based on the clusters. For this purpose, a tweet-wise Deep Neural Network was used [4]. Only clusters that contain at least one crisis-relevant tweet are used for the last step. By forming cluster chains, the clusters of the current interval can be synchronized with the clusters of the previous interval. Similar clusters are merged, new clusters are added and clusters that have not been updated for a long time or clusters that have existed too long are closed.

In order to ensure a robustness towards outliers and to speed up the process, clusters are represented by central centroids (5), which are obtained by averaging a fixed number of cluster representatives that are most similar to each other.

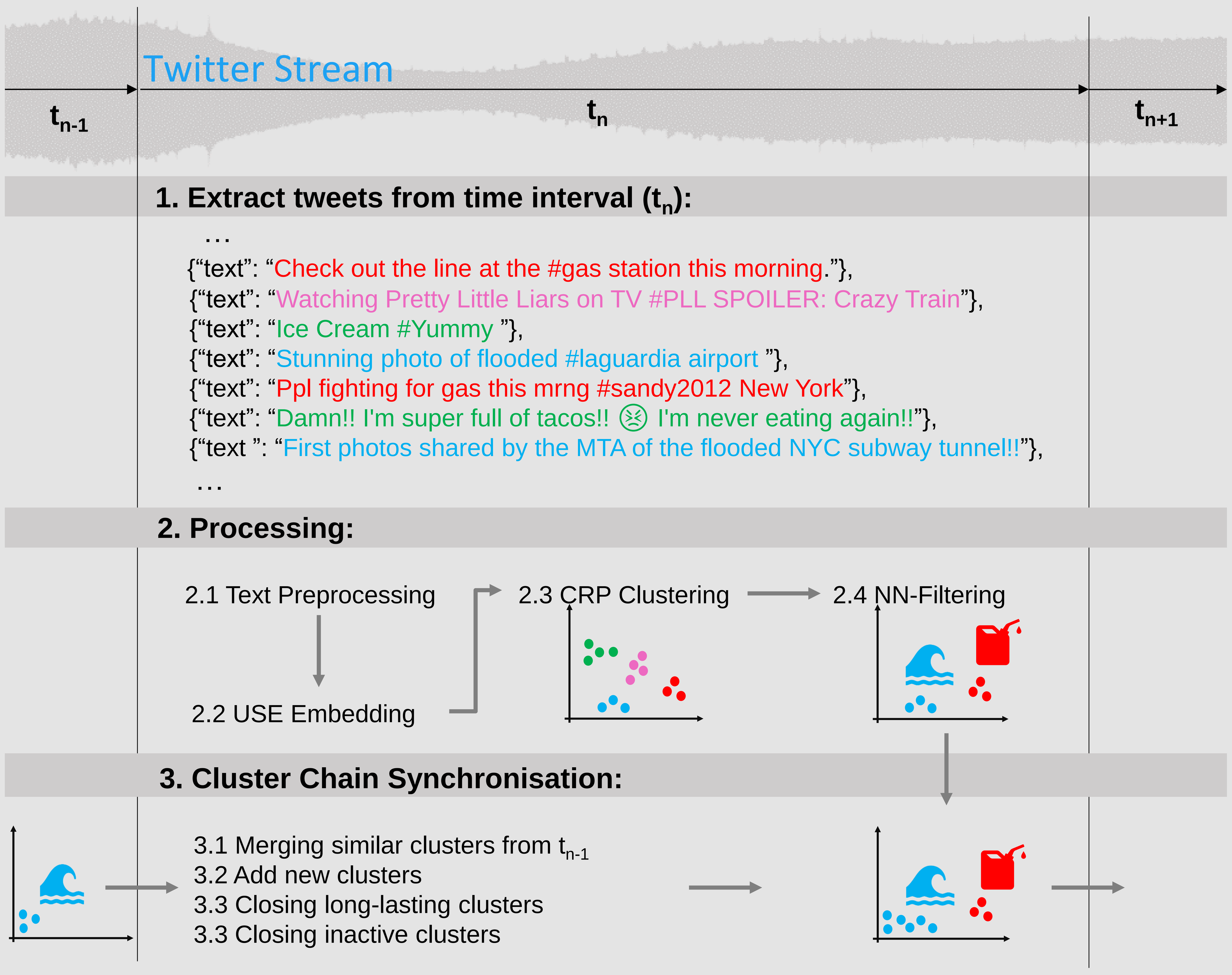

Fig. 2: Workflow dividing the Twitter stream into equal time intervals and filtering the respective tweets on the basis of semantically aggregated clusters

Results:

Since only a few labeled stream datasets are available that have not been thematically filtered, initial analyses were performed using the Event2012 dataset [6].

Initial results show that the hybrid approach can significantly improve the result of pretrained supervised methods. This is especially true for categories in which the supervised model could not be sufficiently pretrained due to missing labels. Furthermore, the clusters contain a high event-related purity, so that parallel events and even subevents can be distinguished [Fig.3]. But further exhaustive quantitative investigations are required in order to minimize the false-positive rate of relevant tweets and thus to push the information overload reduction forward.

The aggregated form of the clusters provides a good starting point for additional analyses such as Part-of-speech-Tagging, spatio-temporal analyses, event or topic detection or content-based summaries. Currently a paper with more detailed information is in the review process [7].

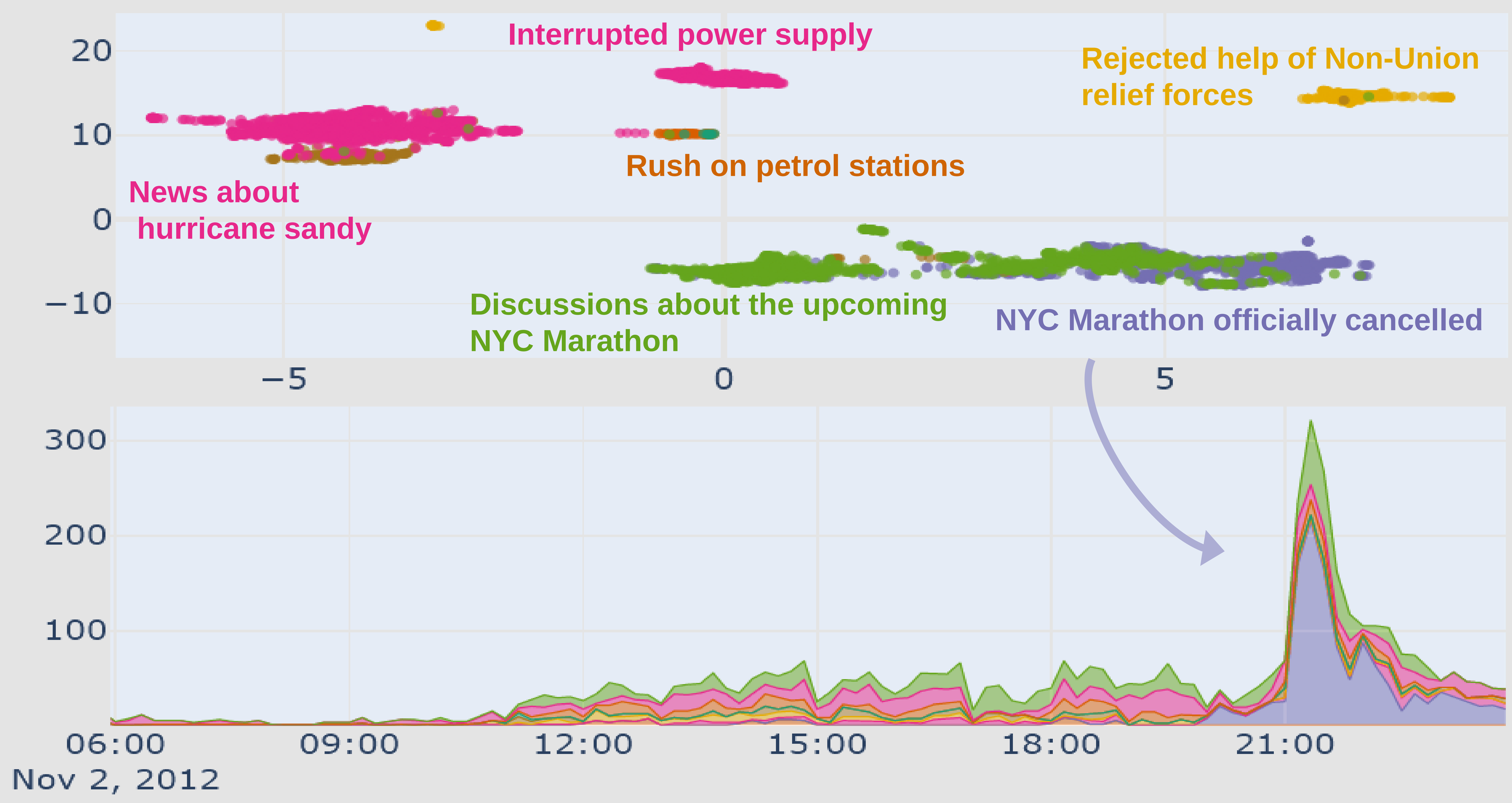

Fig. 3: Illustration from the interactive cluster visualzation tool. Top: 2D UMAP-based visualization of selected clusters containing the keyword „Long Island“. Bottom: Tweet counts over time (GMT) per cluster.

References:
1 Wiegmann, et al.: Opportunities and Risks of Disaster Data from Social Media: A Systematic Review of Incident Information, Nat. Hazards Earth Syst. Sci. Discuss., in press, 2020
2 Kruspe et al.: Review article: Detection of informative tweets in crisis events, Nat. Hazards Earth Syst. Sci. Discuss., in review, 2020
3 Cer, et al.: Universal Sentence Encoder, arXiv preprint arXiv:1803.11175
4 Wiegmann, et al.: Analysis of Filtering Models for Disaster-Related Tweets, ISCRAM 2020
5 Nguyen, et al.: Hot Topic Detection on Twitter Data Streams with Incremental Clustering Using Named Entities and Central Centroids, IEEE-RIVF 2019
6 McMinn, et al.: Building a large-scale corpus for evaluating event detection on twitter, CIKM 2013
7 Kersten, et al.: Combining Supervised and Unsupervised Learning to Detect and Semantically Aggregate Crisis-Related Twitter Content, ISCRAM 2021, in review