RescueMark: Visual Analytics of Social Media Data for Guiding Emergency Response in Disaster Situations

Award for Skillful Integration of Language Model

Astrik Jeitler*   Alpin Türkoglu †   Denis Makarov ‡   Timo Jockers §   Juri Buchmüller ¶

Udo Schlegel ‖   Daniel A. Keim**

University of Konstanz, Germany

ABSTRACT

This paper presents RescueMark, a web-based visual analytics tool for analyzing disaster situations and guiding emergency response. In disaster situations operators must take quick and effective decisions to solve critical problems. RescueMark provides spatial, topic and temporal event exploration supporting decision making for resource allocation and determine damaged areas of the city. We describe the data analysis and visualization process of the social media data applied to extract the relevant information.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Visual analytics.

1 INTRODUCTION

In the VAST challenge 2019 [4], the authors present the fictional city of St. Himark, which is subjected to a disastrous earthquake. The main goal of Mini-Challenge 3 [3], was to analyze the text messages sent from St. Himark citizens using the social media application Y*INT and characterize the conditions in the city to formulate recommendations for the allocation of resources.

The goal was to identify relevant messages related to the disaster and extract needs of the citizens to allocate resources. Several text mining algorithms can be used for this type of problem, such as topic extraction and word embeddings. In a disaster situation it is very important to understand the intention of a message, it's actionability and possible implications for the author and their environment, to understand resource needs and availabilities [5].

In this paper we describe our approach in solving the problem with *RescueMark*, a web-based visual analytics decision dashboard based on the social media messages sent during the disaster situation. The visualization is based on a high-level aggregation of the data which allows it's representations in a clear categorized view and supports the decision making of emergency responders.

2 DATA PRE-PREPROCESSING AND ANALYSIS

In first step we identified important attributes of the data given for the challenge, which contained four attributes: time, location, account name and message. The text messages were very noisy with spelling mistakes and words from languages other than English. Further, the data includes many irrelevant messages such as advertisement. Therefore we defined an analysis pipeline to clean and mine the data to extract relevant messages from the given dataset.

---

*e-mail: astrik-veronika.jeitler@uni-konstanz.de

†e-mail: hafize-alpin.tuerkoglu@uni-konstanz.de

‡e-mail: denis.makarov@uni-konstanz.de

§e-mail: timo.jockers@uni-konstanz.de

¶e-mail: juri.buchmueller@uni-konstanz.de

‖e-mail: u.schlegel@uni-konstanz.de

**e-mail: keim@uni-konstanz.de

2.1 Extracting Crisis-Relevant Data

First, we used a spell-checking algorithm to correct out-of-dictionary words contained in many messages.

In the next step, we filtered not relevant data, mainly advertisement messages, from the dataset. This step was necessary as many of those messages contained misleading homonyms for keywords related to the disaster. An example can be found in the message *"appealing, trembling. What better deals can you ask for?"*. Here, it would be misleading to falsely attribute the word *"trembling"* to the earthquake. To minimize such distortions, we transformed the messages into word embeddings using the fastText [2] algorithm and applied *k*-means clustering to remove clusters of unrelated messages from the dataset.

2.2 Characterizing Conditions and Resource Allocation

To characterize conditions and allocate resources, we divided the cleaned data into messages related to the disaster and messages not related to the disaster. Messages related to the crisis were classified into 10 categories. For this, we applied the Latent Dirichlet Allocation (LDA) algorithm [1] to extract topic terms for each messages. The 24 highest weighted topic terms from the top 20 topics were selected as query words for message retrieval in the fastText vector space. The messages were thus classified according to 10 resources classes. The final step was to extract messages only expressing a "need" for a certain resource. This was achieved using both pattern-matching and manual annotation.

We focused on identifying exact earthquake times in order to detect fatalities, damages, and conditions throughout the city. Using the subset of disaster relevant messages, we identified three earthquakes by finding the cluster of messages related to the topic *"Earthquake"*. On this small dataset we performed a burst analysis using Kleinberg's burst detection algorithm to find times when an increase of messages mentioning an earthquake could be found [6]. These times then were used in a to find the first most reliable reports of earthquakes.

2.3 Pattern-matching for Fatalities Reports

The fatalities were detected through clustering of the messages after transformation into word embeddings using the fastText algorithm. From the respective cluster, we extracted the number of fatalities by analyzing the textual patterns. We found five different patterns in the "fatalities" dataset that were use to determine the number of fatalities and the certainty of the report.

3 VISUAL ANALYTICS

RescueMark presents all relevant information in a single dashboard (see Figure 1). The interaction possibilities at the users disposal are for the singular purpose of "zooming" into the data to focus on certain data points, such as a specific time interval or district. The dashboard is divided into 5 sections: Social Navigation Bar, Interactive City Map, Events, Timeline, and Need for Resources - Trend Topics.
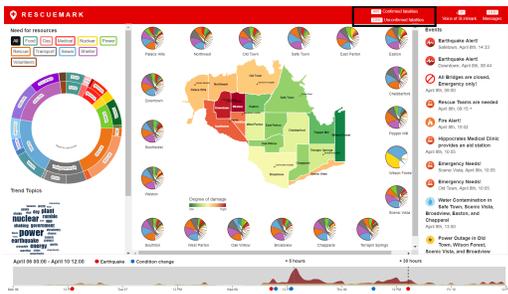
Figure 1: RescueMark - dashboard overview

3.1 Interactive City Map

The interactive map of St. Himark shows the degree of damage per district in a linear scale (see Figure 2). Each district of the city is selectable, to update the surrounding charts and information based on the region in a "District View". Users can detect damaged areas after the earthquake according to the chosen time range.
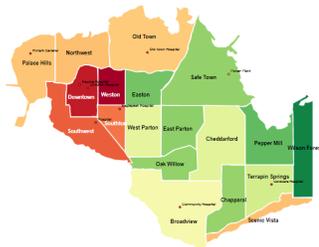

Figure 2: Interactive map: degree of damage per district

The chord charts around the map show the resource needs for each district at the selected time interval. Each color in the is encoded with the related resource across the whole dashboard.

3.2 Need for Resources and Trend Topics

The map can be changed to show the need for a specific resource for all district by selecting a specific resource in the resource menu (see Figure 3).

Figure 3: Resource menu

Operators can choose a specific resource with sunburst diagram to compare the need for the selected resource between all districts using the color scale (see Figure 4).

To get the pulse of the community, we wanted to use word clouds to show trend topics based on social media data. There are two types of word clouds in the tool, the first one shows the most frequent words of users, and the second one shows the most frequently used hashtags. With these charts, operators can get insights about the city.
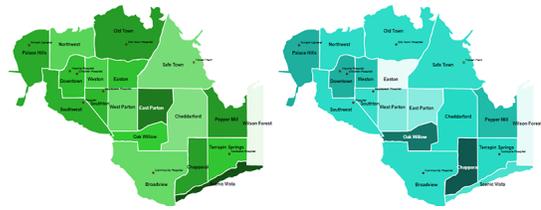

Figure 4: Need for the resource views on the city map *Power* (left) and *Food* (right)

3.3 Events and Decision Maker Announcements

The "Events" list includes the critical situations an official announcement in St. Himark. These events can be to select the corresponding time interval and quickly characterize conditions for this event.

3.4 Social Navigation Bar - The Voice of St. Himark

The *Fatalities* section shows the uncertainty or certainty of the data in a specific time interval. This feature helps direct operators in understanding how to dispatch rescue and medical assistance teams.

Figure 5: Social Navigation Bar

The subset of messages shown are divided into two sections: messages related to social activities during the disaster to reflect the *Voice of St. Himark* and messages related to the crisis.

4 CONCLUSION

The decision dashboard RescueMark includes resource needs, decision marker announcements and critical events as well as fatality reports in one overview and offers user-interaction features to focus on specific spatio-temporal units using a time slider and interactive map. We described how a combination of data mining and visual analytics techniques could be applied to solve the challenge. We showed how word embeddings and clustering can be applied to extract relevant crisis related resource needs related messages from a noisy dataset and event detection could be used to find changes in conditions for resource allocation. Moreover we demonstrated that an efficient combination of pattern-matching analysis and contextual information could be helpful to extract reliable text messages from fatality reports.

REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[2] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*, 2016.

[3] K. Cook, J. Fallon, and J. Crouser. Mini-Challenge 3: Voice from the People. https://vast-challenge.github.io/2019/MC3.html, 2019.

[4] K. Cook, J. Fallon, and J. Crouser. VAST Challenge 2019: Disaster at St. Himark! https://vast-challenge.github.io/2019/index.html, 2019.

[5] M. Basu et al. Automatic matching of resource needs and availabilities in microblogs for post-disaster relief. In *Companion Proceedings of the The Web Conference 2018*, 2018. doi: 10.1145/3184558.3186911

[6] W. Dou et al. Event detection in social media data. In *IEEE VisWeek Workshop on Interactive Visual Text Analytics-Task Driven Analytics of Social Media Content*, pp. 971–980, 2012.