COMP4462 Project Report
VAST Challenge 2019
Mini Challenge 1: Crowdsourcing for Situational Awareness

Fei NIE	**Yao ZHANG**	**Huan Ting WEI**	**Yankun ZHAO**
20583541	20583319	20469903	20583371
Contribution: 25%	Contribution: 25%	Contribution: 25%	Contribution: 25%

Abstract

Our Earthquake Damage Report Interactive Dashboard addresses Mini Challenge 1 of VAST Challenge 2019. The dashboard utilizes interactive visual analytics to report and summarize crowd-sourced damage reports temporally and spatially. We utilized a relative standard deviation to evaluate the uncertainty of data. On the aspect of priority, Sum of Inverse-Variance Weighted Intensity Values was implemented. Definition and detailed description of the algorithms will be illustrated later. Data pre-processing was done in Jupyter Notebook with Python, and the interactive visualization was implemented with D3 and React. The dashboard can be found at https://alicenie.github.io/comp4462_project/.

Keywords: Human-centered computing, Visualization, Visualization application domains, Visual analytics

1 Introduction

VAST Challenge 2019 Mini Challenge 1 gave a fictitious situation: Earthquakes hit a city called St. Himark. With a mobile application, its citizen can report the shake intensity and damage on power, medical, buildings, sewer and water, and roads and bridges on a scale of 0 to 10. Using the data provided, our visualization aims at answering the following questions to assist the emergency responders in St. Himark:

- How should emergency responders respond based on the damage reports from citizens?
- How to prioritize which neighborhoods to rescue first?
- When and where is the major earthquake hit?
- How reliable are the reports?

2 Data Processing

2.1 Data Preprocessing

There are `83,070` data entries in the dataset, and each record in the original dataset describes a single damage reporting event recorded for every 5 minutes, including the timestamp, location and damage severity (rating from 0 to 10) of earthquake shake intensity, power, buildings, medical, roads and bridges, and sewer and water. Missing data points were replaced by `-1`. Also, as the original time interval was relatively small and unsuitable for visualization, we changed the scaling to 30 minutes by gathering all records within every 30 minutes. We then used the preprocessed data to calculate the uncertainty and priority of the responses, which will then be visualized in the dashboard to help resources allocation during the emergency. The statistical intermediate data including count, mean, variance, standard deviation, uncertainty, and priority were all pre-calculated before being consumed by APIs.

2.2 Calculation of Uncertainty

Uncertainty of data has a vital impact on the reliability of data, without which data cannot be used by rescue teams to make decisions. Also, assume that emergency responders are always required to pick a region to rescue, only relative uncertainty/certainty is critical in this case. Below are the formulas for computing the uncertainty:

$$SD = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \overline{x})^2}{n-1}}$$

$$scaler = \frac{SD - SD_{min}}{SD_{max} - SD_{min}}$$

$$SD_{scaler} = SD_{min} + scaler \cdot (SD_{max} - SD_{min})$$

We use statistical dispersion to measure data uncertainty. We calculate the standard deviation of data points with the same attribute during each time interval (which is 30 minutes) for different regions, we then apply a scaler on it to make sure that it can return a number in a range of 0 to 1, which is convenient for visualization using a color palette. And here, we chose to use standard deviation to measure the dispersion of the data points, and the relativity is shown among the different regions per attribute per time interval. We also considered using normalized entropy to measure uncertainty. However, with normalized entropy, a case with three 1, three 10 and a case with three 5, three 6 would give the same result, which is not ideally what we want.

2.3 Calculation of Rescue Priority

To calculate the rescue priority of one region based on both the severity and reliability of the reported situation, we conducted a literature review and got inspiration from some previous studies [1, 2]. We proposed the Sum of Inverse-Variance Weighted Intensity Values as the measure of Rescue Priority. Given a region i and a 30-minute time interval, the rescue priority value of this region during the interval is defined as the formula below:

$$RPV_i = \sum_{d \in D} \frac{\mu_R}{\exp(\lambda \times \sqrt{\frac{\sum_{r \in R}(r - \mu_R)^2}{|R|}})}$$

It considers all the 6 kinds of damage information, and each of them is weighted by its inverse variance. Additionally, we use λ as a hyperparameter to control the significance of data uncertainty to the rescue priority, which can be tuned depending on the application. For example, if the rescue power is very abundant or if all the damage information is sufficiently credible, then λ can be set to a very small number, which means not caring about the reliability or uncertainty of the reported damages. In our case, we use $\lambda = 1$. Moreover, the exponential function in the denominator is used to avoid division by zero and to assign fewer weights to highly dispersed reported damages.

3 Visualization Design

To provide a visualization with "**overview, zoom and filter, and details on demand**" for our target users (i.e., the emergency response teams in St. Himark), we created a dashboard that contains 4 views:

- **City Map** a selectable choropleth map of St.Himark city showing the rescue priority of each neighborhood
- **Neighborhood Breakdown** a radar chart showing the breakdown information of the selected neighborhood
- **Earthquake Hit Timeline** two bar charts showing the number of reports and shake intensity over the whole earthquake period, with synchronized draggable time span
- **Damage per Neighborhood and Service** a customized heatmap with two color palettes

The time span of all visualizations is synchronized using the draggable time span in Earthquake Hit Timeline.

3.1 City Map and Neighborhood Breakdown

City Map is a choropleth that uses color to show the priority of each area. The color luminance encodes the rescue priority, where darker color indicates higher priority. The map also highlights the locations of the critical facilities such as nuclear plants and hospitals by overlaying corresponding symbols.

As this is the first graph the emergency response team will see when entering our interface, we aim to provide a **most high-level overview** that gives the audience a

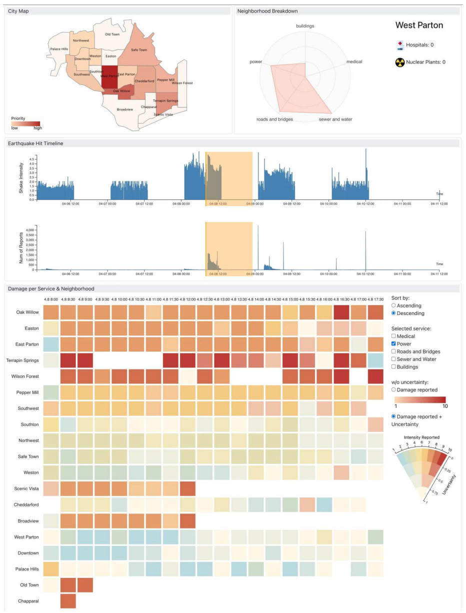

Figure 1: Dashboard

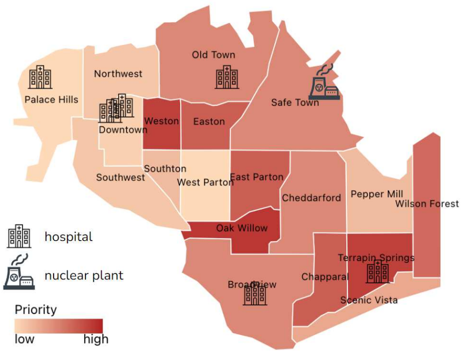

Figure 2: City Map of St.Himark

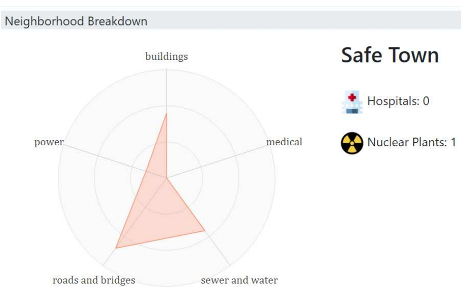

Figure 3: Neighborhood Breakdown - Safe Town

first-glance intuition, allowing them to make decisions such as where to go and what to do first swiftly. Therefore, City Map does not include complicated raw, numerical data, instead, it presents the priority with the simplest encoding scheme.

That being said, **raw data that constitutes the calculation result can still be viewed** by selecting any of the neighborhoods. When clicking on a neighborhood, the reported damage intensity of all service types will be displayed in a radar chart juxtaposed with City Map. The five service types, *buildings*, *medical*, *sewer and water*, *roads and bridges*, and *power*, are evenly distributed in the radar chart. The radius lengths represent the value of the reported damage, which is on a scale of *0 to 10*. This breakdown also displays the number of critical facilities located in this neighborhood.

3.2 Earthquake Hit Timeline

Figure 4: Earthquake Hit Timeline

Earthquake Hit Timeline includes two bar charts that display the number of reports received and the shake intensity overtime respectively. The yellow area shown on both charts is the time span applied globally to the whole interface. Users can change it by dragging it. There is also a tooltip on the hovered time that shows the actual value.

The purpose of the timeline is beyond changing the time span of the interface. Moreover, it provides multiple valuable insights. The time of the main hits can be easily observed from the matching spikes of the two charts as well as the smaller hits with fewer affected areas. A large number of delayed reports is also revealed by the comparison of the two charts, which might indicate the need for more timely responses.

3.3 Damage per Service and Neighborhood

As there are rescue teams consisting of different expertise in the 5 service types, it is crucial that our visualization **filter out the unneeded information and only display the relevant parts**. Therefore, we separated the damage of each service type and display them with a heatmap. The 5 service types are only shown once it is selected using the selection panel on the right. The y-axis lists out the names of the neighborhood and the x-axis shows the time. The ascending and descending sorting option can also help users to find the neighborhood with the highest or lowest damage.

There are two different color schemes. One is a **normal linear color scale** that encodes the damage reported on a scale from 0 to 10. The other is the **Value Suppressing Uncertainty Palettes(VSUP)**[3] that has an extra dimension representing the uncertainty level calculated earlier. The lighter the color is, the more uncertain the reported value is.

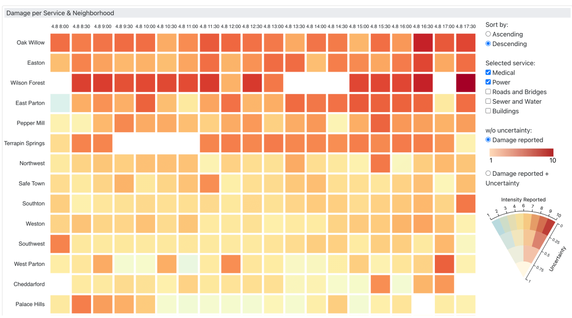

Figure 5: Damage Reported

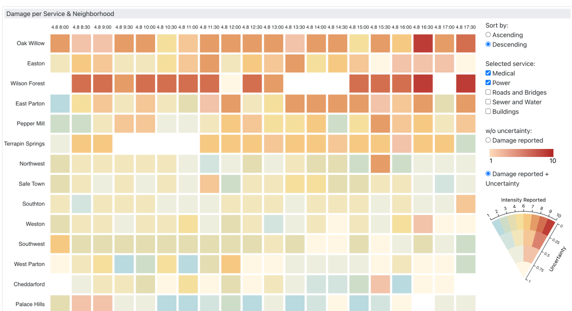

Figure 6: Damage Reported + Uncertainty

The selection panel for different services helps the rescue team to **focus on their expertise without distraction**. The medical rescue team, for example, may only want to see the damage of medical facilities while the engineering team may want to see roads and bridges. It is also a clear way to **compare the damage of different services** in the neighborhood.

4 Conclusion

This paper presents an integrated dashboard that addresses the problem of emergency resource prioritization and allocation after earthquake hits described in VAST Challenge 2019 Mini Challenge 1. The dashboard consists of 4 visualization components, which are City Map, Neighborhood Breakdown, Earthquake Hit Timeline, and Damage per Service and Neighborhood. The integration of the 4 visualization components provides a first-glance overview as well as detailed insights for the emergency responders to make both immediate and well-thought-out decisions. The measure of uncertainty also highlights the reliability of the received data, rendering a more trustworthy and accurate result.

There are some works that could be done in the future. First, since the current time range is set to be 10 hours due to the limited width of the heatmap, we suggest allowing the user to adjust the length of the time span. Second, as an interactive interface, user studies could be done to evaluate the usability and usefulness of different visualization designs.

References

[1] F. Marin–Martinez en J. Sanchez–Meca, 2009. "Weighting by Inverse Variance or by Sample Size in Random-Effects Meta-Analysis", in *Educational and Psychological Measurement*, 70(1), pp.56-73.

[2] Y. Jiao, M. Cornec, en J. Jakubowicz, "An entropy-based term weighting scheme and its application in e-commerce search engines", in *International symposium on web algorithms*, 2015.

[3] M. Correll, D. Moritz, and J. Heer. 2018. "Value-Suppressing Uncertainty Palettes. Proceedings of the 2018

CHI Conference on Human Factors in Computing Systems", in *Association for Computing Machinery*, New York, NY, USA, Paper 642, 1—11. DOI:https://doi.org/10.1145/3173574.3174216