VAST Challenge 2019 Mini-Challenge 1: Damage Reports and Visual Analysis of the Earthquake in St.Himark

Ashini Anantharaman Niranjana Sathish Avilery Saitulasi Kamma Mithila Nigudkar
Department of Software Engineering
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, United States

***Abstract*—This paper elucidates the implementation of the Vast Challenge 2019 Mini Challenge 1 (MC1) which is based on the town St.Himark which has been hit by an earthquake. In this study, visual analytics techniques are used to summarize the crowd-sourced damage reports temporally and spatially to analyze disaster situations and guide emergency response. Visualizations were also done to understand uncertainty and reliability in the data obtained. Tableau, Google Charts, and Python libraries were the tools used in the project.**

I. Introduction

The town St.Himark has been hit by an earthquake, forcing officials to determine the extent of the damage. In response to this event, the teams receive seismic readings and tried to use them for initial deployment. However, more information was required to completely analyze the true conditions throughout the city. Prior to the earthquake, the city had released a mobile application for reporting the intensity of the damage. To help authorities focus on recovery efforts, citizens are playing a crucial role by providing them with needed information about the effects of the earthquake. Mini-Challenge 1 data includes one CSV file that contains (categorical) individual reports spanning the entire event, showing the extent of shaking/damage to different neighborhoods. The data was collected from an application that allows citizens of this city to upload damage information for six different categories, such as buildings, roads, bridges, power, health conditions, sewers and water, and finally, shake intensity. People respond to this app by ranking the conditions for each of the attributes for which they have a preference from 1 to 10. [1]

II. Data Exploration

In this challenge, the main data consisted of 83,071 lines, each representing a user-submitted report that featured time, location, and rating for each category of damage. A python script was compiled to determine the total number of missing values in various features and also to generate a five point summary which includes mean, standard deviation, and other statistical parameters to understand more about the distribution of data.

- To get a vivid insight of the above statistics, a box plot was used on each of the categories which resulted in an outlier for the shake intensity that was caused due to a low average and an exception of few high ratings.
- The medical category had 57% of its data missing followed by shake intensity with 14.6% which was determined using a python library named Seaborn.
- To get a profound insight on the data distribution, histogram was used which resulted in the highest count being 4485 reports on 9th April, 2020 at 01:00 p.m. and 3,902 reports on 10th April, 2020 at 12:00 p.m being the second highest. Even though the first and the second highest reports were on 9th and 10th, it was inferred from the graph that day 8 had a lot of reports over different times of the day.
- The count of reports with respect to different locations in St. Himark were visualized using bar charts that resulted in Scenic Vista and Old Town having the most number of ratings with more than 13,000 values whereas the region Wilson Forest had only 173 reports.
- Finally, to get a precise report on each of the categories, a python library named Sweetviz was used that exhibited a bar chart from which it was inferred that in most of the categories, data distribution was left-skewed with an exception of the building ratings which was distributed normally.

III. Visualization and Analysis

Based on the previous exploratory analysis of data, a relative study was implemented between two regions which has the most(Scenic Vista) and least(Wilson Forest) number of ratings from the citizens. The bubble chart in Figure 1 distinctly represents the days and hours at which inputs were received and the size depicting the average value of ratings. To analyze which neighborhood was the worst-hit and to prioritize the region that needed the most attention over a period of time, Tableau and the shape file provided in the Mini-Challenge 2 were used. Figure 2 depicts a geographical heatmap [2] based on the average ratings provided for each feature which was filtered for a certain day and hour (Day

Fig. 1. Comparison of Scenic Vista and Wilson Forest using Bubble chart

8, hour 8). It can be derived that places like Old Town, Northwest and downtown are affected in all the categories for that particular time. Similarities were found in all the groups with the exception of shake intensity. This can be seen in the locations like Palace hills, Scenic Vista and BroadView.

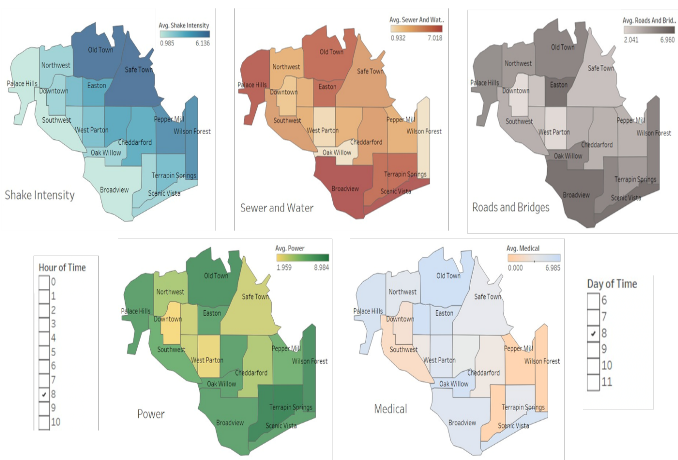

Fig. 2. Spatial distribution of data on Day 8 at 8:00AM

To further verify the uncertainty in shake intensity, a comparative line graph was created between these features as shown in Figure 3 with respect to location. This proved the previous hypothesis that the shake intensity values were very different from other categories. In regions like Broadview and Chapparal, it was observed in certain instances that when the shake intensity values were low, other values were high.

Based on the previous observations, a specific instance was identified which substantiates the unreliability of date in the shake intensity class. Figure 4 provides the ratings for this group on a particular day in three neighboring places namely Old Town, Safe Town and Easton in a tablular format. On April 9, 2020 at 12 AM, the ratings provided by the public were ranged from 0 to 8 which clearly depicts an anomaly.

IV. Conclusion and Future Work

Based on the above analysis, at specific time period it was possible to determine which area was most impacted by the earthquake for these categories but it was also evident

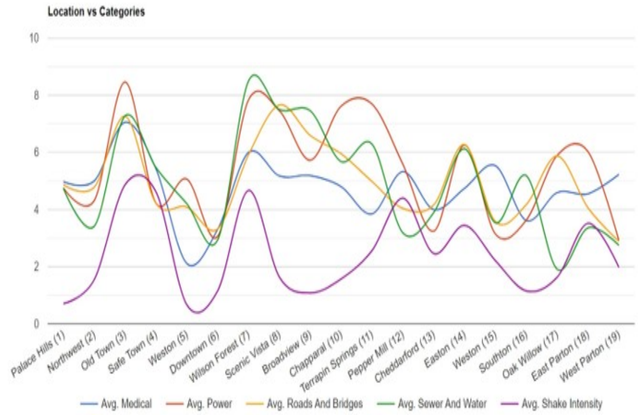

Fig. 3. Average measure in each category vs Location

Hour of Time	Nbrhood / Id: Easton (14)	Old Town (3)	Safe Town (4)
9 April 2020 12 AM	1		
9 April 2020 1 AM	1	0 1 2 3 4 5 6 7 8	
9 April 2020 2 AM	1	0	
9 April 2020 3 AM			
9 April 2020 4 AM	0 1	0 1	
9 April 2020 5 AM		0	
9 April 2020 6 AM	1	0	0
9 April 2020 7 AM			
9 April 2020 8 AM			1
9 April 2020 9 AM	1	1	
9 April 2020 10 AM		0 1	0
9 April 2020 11 AM		0	
9 April 2020 12 PM	0 1	1	0
9 April 2020 1 PM	1	0	
9 April 2020 2 PM		0 1	
9 April 2020 3 PM	1 2 3 4 5	3 4 5 6 7	3 4 5 6 7
9 April 2020 4 PM	1 2 3 4 5		1 3 4 5 6 7
9 April 2020 5 PM	0 1 2 3 4 5		3 4 5 6 7
9 April 2020 6 PM	2 3 4 5		4 5 6 7
9 April 2020 7 PM	0 2 3 4		1 3 4 5 6
9 April 2020 8 PM	2 3 4		3 4 5 6
9 April 2020 9 PM	0 2 3		4 5 6
9 April 2020 10 PM	0 1 2 3		4 6
9 April 2020 11 PM			1

Fig. 4. Rating comparisons of three neighboring locations

that there can be instances of uncertainty in the data. To enhance the geographical analysis, tools like QGIS can be used to create a time temporal for the entire time period of this earthquake in a real time basis with a heatmap that depicts an alert based on the ratings provided by the citizens. Also, models like Bayesian time series can be use for efficient statistical analysis of the given features which will be reliable and not time consuming.

References

[1] A. Majumdar, G. Ymeri, S. Strumbelj, J. Buchmüller, U. Schlegel, and D. A. Keim, "Earthquake investigation and visual cognizance of multivariate temporal tabular data using machine learning," in *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2019, pp. 136–137.

[2] A. Jeitler, A. Türkoglu, D. Makarov, T. Jockers, J. Buchmüller, U. Schlegel, and D. A. Keim, "Rescuemark: Visual analytics of social media data for guiding emergency response in disaster situations."