# Interactive Ranking Uncertain Multivariate Ordinal Time Series

Shichao Jia* Jiaqi Wang Zeyu Li Jiawan Zhang

Tianjin University

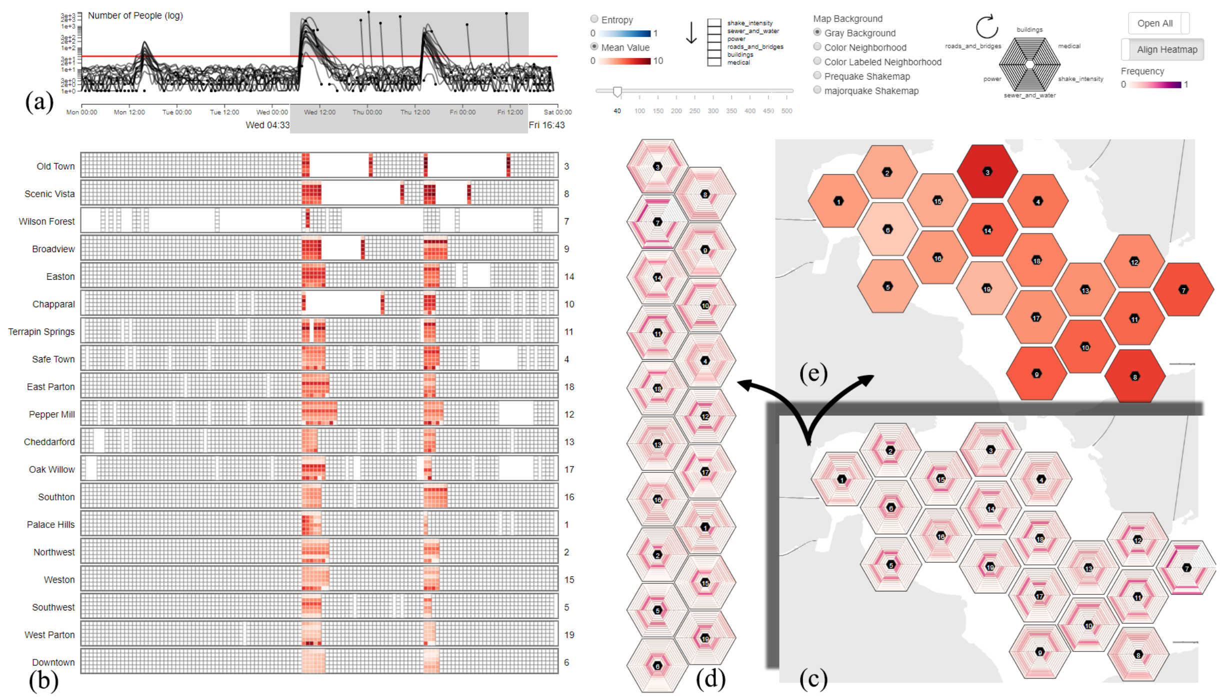


Figure 1: The interface of our system. (a) The line chart summarizes the number of people who upload records of each neighborhood. (b) The heat map visualizes mean value or entropy along the time for each neighborhood of different dimensions. (c)-(e) Hexagons on the map can be aligned with the heatmap.

## Abstract

This work proposes a visual analytic technique to interactive rank uncertain multivariate ordinal time series for situational awareness. We use normalized entropy to quantify the uncertainty of the discrete survey data. The heatmap summarizes both the survey data and qualified uncertainty in a compact manner. Besides, the distributions of votes in neighborhoods are visualized using hexagons. This work presents the result of applying the solution to the VAST 2019 - Mini-Challenge 1 dataset, which led to the Award for Excellent Quantification of Abnormalities.

**Index Terms:** Human-centered computing—Visualization—Visual analytics

## 1 Introduction

Mini Challenge 1 (MC1) of the VAST Challenge 2019 aims to prioritize neighborhoods for the response after earthquake considering both emergency and uncertainty. The data comes from apps of nineteen neighborhoods, collecting different levels of different damages along the time. Therefore, the data can be viewed as multivariate uncertain ordinal time series for different neighborhoods. The key challenges of MC1 include how to quantify the uncertainty, how to visualize the temporal uncertainty and how to rank the neighborhoods considering different damages each time.

Facing these challenges mentioned above, we propose a visual analytics system to help users interactive rank the uncertain multivariate ordinal time series. We first propose to use normalized information entropy [1] to measure the uncertainty, because it is suitable to quantify the reliability of ordinal survey data. Then, we use the heatmap to encode the mean value and the entropy for different neighborhoods of different damage type along the time. Besides, we provide several interactions to enable users to rank the time series under different circumstances. Using our tool, we are able to quickly identify which regions provide more reliable reports and which regions are more urgent to rescue. Besides, we can detect the data delay phenomenon due to the power outages.

## 2 Data and Processing

MC1 provides a dataset spanning the entire length of the earthquakes, containing categorical reports of shaking/damage by neighborhood over time (every five minutes). There are six dimensions including shake intensity, sewer & water, power, roads & bridges, medical and buildings. Each dimension has eleven levels (0 for the lowest and 10 for the highest).

For each dimension of each timespan, we use mean value $S(d_i)$ to measure the damage ($d_i$ for each dimension). For uncertainty, we

*e-mail: jsc_se@tju.edu.cn

2019 IEEE Conference on VAST
20-25 October, Vancouver, BC, Canada
978-1-7281-2284-7/19/$31.00 ©2019 IEEE

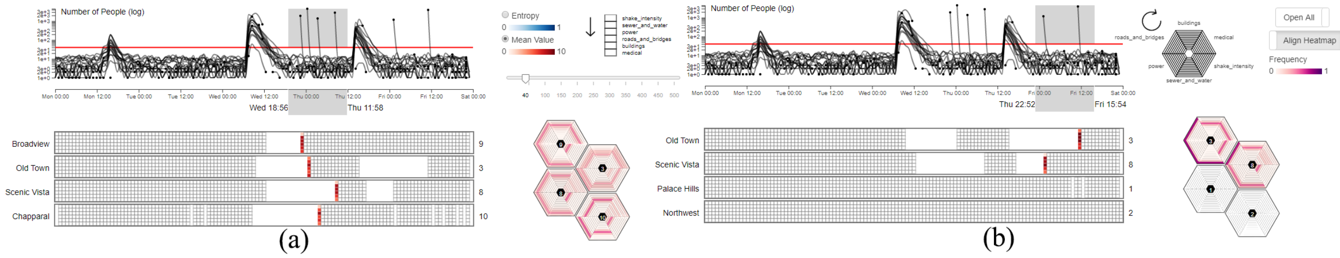


Figure 2: Select the breakpoints after the two earthquakes. We can find that the data missing is due to the power outages.

choose normalized entropy because it is more suitable for ordinal survey data:

$$H_{11}(d_j) = -\sum_{i=0}^{10} p_i(d_j) log_{11} p_i(d_j)$$

in which $d_j$ stands for one dimension, $p_i(d_j)$ is the frequency of the level $i$. We choose base 11 (total 11 measurement levels from 0 to 10) to normalize information entropy so that the measurements will range from 0 to 1. The higher the entropy is, the more uncertain the reports are, which means citizens have less consistent reports.

To rank the time series based on the emergency and reliability of each neighborhood, we propose to use the following simple but intuitive metrics for measurements:

$$\overline{S} = \frac{\sum_{d_j}\sum_{t_k} S(d_j)/|t_k|}{|d_j|}, \overline{H} = \frac{\sum_{d_j}\sum_{t_k} H_{11}(d_j)/|t_k|}{|d_j|},$$

in which $\overline{S}$ measures the overall damages considering selected dimensions in the selected time ranges, and $\overline{H}$ measures the mean entropy for reliability.

## 3 Visual Design

After several discussions inside our team, we conclude three-level visualization tasks based on the abstraction levels of data.

- (T1) Rank neighborhoods directly based on the emergency or reliability of reports.
- (T2) Compare the damage levels or uncertainty between different dimensions.
- (T3) Visualize vote distribution of citizens of each dimension of different neighborhoods.

These three-level tasks abstract data from coarse to fine-grained. Visualizations should quickly respond due to the urgency. Therefore, T1 directly concludes which neighborhoods should be first rescued from a high level. Then, emergency responders want to check what resources should be allocated to these neighborhoods. T2 decompose the data into different dimensions for easy comparison. Furthermore, if the time permits, responders may want to check how the votes distribute from different levels to verify their decisions. T3 will provide more details at a low level for a closer investigation.

Based on these considerations, we propose two kinds of visualizations including hexagons and heatmaps. For T3, each hexagon encodes the vote distributions on each dimension of a neighborhood. Each pan stands for one dimension. Levels layout from inner to outer. The color stands for the frequency of each level. This visualization provides how reliable the reports are and which dimensions we should pay more attention. For example, unreliable reports distribute evenly because citizens have divergent votes. However, reliable reports distribute centralized on specific levels. Therefore, responders should focus on the centralized reports with data distributing at the outer ring. We provide two kinds of layouts of the hexagons. Hexagons can locate based on relative spatial locations of neighborhoods (Fig.(c)). Although this layout preserves the spatial information, it makes the ranking unclear. Therefore, we provide the second method to layout hexagons based on the ranking order to align with the heatmap (Fig.(d)).

For T2, a heatmap visualizes the time series in a compact manner. Each row stands for one neighborhood, and it is decomposed into different dimensions. Cell color encodes mean value or the normalized entropy (white cell without black border stands for missing data). For T1, the heatmap will be ordered based on the measurements $\overline{S}$ and $\overline{H}$. Besides, when the hexagons align with the heatmaps, the color of hexagons on the map will directly encode $\overline{S}$ and $\overline{H}$ so that both the ranking and spatial information are preserved (Fig.(e)). Meanwhile, we provide several interactions to rank the time series. A line chart indicates users which regions of data should be focused on (Fig.(a)). Each line stands for the number of reports of each neighborhood. Users can select the data over a threshold of people in the selected time range.

## 4 Analyzing VAST 2019 - MC1 Dataset

We first find that there are three peaks along the time in the line chart (Fig.(a)). Besides, there are several breakpoints which may be the delays of the reports due to the power outages after each earthquake. Besides, After switching to the entropy, the heatmap shows that in the most time, whenever the reports increase, the uncertainty decrease (see our submission file). Therefore, we should pay more attention to the reports with more people, since these reports are more reliable. Besides, we can choose an appropriate threshold to filter out unreliable reports. We choose the last two events to rank the neighborhoods. The result is shown in Fig..

Furthermore, we can choose only breakpoints after the first earthquake (Fig.2). We notice that reports of Broadview, Old Town, Scenic Vista, and Chapparal are missing. But after 12-20 hours, the reports increase explosively. Therefore, we infer that it is due to the power outages of these neighborhoods. Besides, we can find all their power are urgent for the emergency, which confirms our hypothesis. A similar pattern can be found after the second earthquake (Fig.2).

## 5 Conclusion

This work proposes a visual analytics solution to help users interactive rank uncertain multivariate ordinal time series. We propose three-level visualizations based on derived tasks using both heatmap and hexagons. The application of this solution to VAST 2019-MC1 dataset confirms our design decisions.

## References

[1] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.