Visualizing Uncertainty with UofC-Bayes, Mini-Challenge 1

Haysn Hornbeck*
University of Calgary

Usman Alim†
University of Calgary

Abstract

Disasters demand a quick response based on incomplete information. For the Saint Himark dataset, part of the 2019 VAST Challenge, we focused on delivering a visualization which accurately conveyed that uncertainty. While our analysis was done offline, we chose techniques and algorithms which could easily be applied to real-time usage. Our visualization for the first mini-challenge was a one-screen dashboard that summarized citizen feedback.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Graph drawings; Computing methodologies—Modeling and simulation—Simulation types and techniques—Uncertainty quantification; Applied computing—Physical sciences and engineering—Mathematics and statistics—

1 Introduction

Our entry dealt with citizen feedback on earthquake intensity and damage, collected via the RUMBLE smartphone app. This consists of the district they are in, the current earthquake intensity on a Likert scale ranging from zero to ten, and an assessment of the current damage for five types of infrastructure on the same scale.

1.1 Theoretical Model

When no earthquake has occurred, citizens will periodically report earthquake intensities of zero or one and randomly choose a damage level, with a slight preference for reporting the maximal damage possible. We can exploit this behaviour to model each district's nominal packet rate via the Poisson distribution and Gamma conjugate prior, both of which are commonly used to represent rates. [1] Updating a conjugate prior is computationally trivial, allowing real-time usage.

At each time interval, we could estimate the likelihood of observing the number of reports via the posterior predictive distribution,

$$p(c \mid \alpha,\beta) = \mathrm{NegBinomial}(c \mid \alpha, \frac{1}{1+\beta}) \tag{1}$$

$$= \binom{c+\alpha-1}{c}(1-\frac{1}{1+\beta})^{\alpha}(\frac{1}{1+\beta})^{c}, \tag{2}$$

where c is the packet count, and α and β are the Gamma prior's hyperparameters.

This conjugate prior allows us to determine if the number of reports arriving in a specific timespan are abnormally high, indicating an earthquake event, or abnormally low, indicating a power outage. As a prior, we used $\alpha = \frac{1}{3}, \beta = 0$, due to its ability to approximate the maximal likelihood. [4] We found the best procedure was to normalize the posterior predictive of the observed count by the maximal likelihood count. After performing cancellations and converting factorials to gamma functions, to allow for fractional modes, the normalized metric became

$$\mathrm{Q}(c \mid \alpha,\beta) = \frac{\Gamma(c+\alpha)\Gamma(m+1)}{\Gamma(m+\alpha)\Gamma(c+1)}\left(\frac{1}{1+\beta}\right)^{c-m},$$
$$m = \begin{cases} \frac{\alpha-1}{\beta}, & \alpha > 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

This metric ranged between zero and one, and was easily converted to a more intuitive one which ranged between negative one and one; positive values indicate more reports than expected over the given timespan, while negative values indicate fewer.

$$\mathrm{L}(c \mid \alpha,\beta) = \begin{cases} 1-\mathrm{Q}(c \mid \alpha,\beta), & c > m \\ \mathrm{Q}(c \mid \alpha,\beta) - 1, & \text{otherwise.} \end{cases} \tag{4}$$

Damage reports per infrastructure type were modelled via the multinomial distribution and the Dirichlet conjugate prior, as these could represent multimodal data. [1] Each category follows a Beta distribution, specifically

$$p(k \mid I,D) \sim \mathrm{Beta}(\alpha_{I,D,k}, -\alpha_{I,D,k} + \sum_k \alpha_{I,D,k}), \tag{5}$$

where $\alpha_{I,D,k}$ is the Dirichlet hyperparameter for district D, infrastructure type I, and Likert category k. $p(k \mid I,D)$ represents the probability distribution of the odds of a given category being chosen for a specific district and infrastructure type.

The Dirichlets were stored as cumulative sums, as this allowed $O(1)$ calculation of the hyperparameters over an interval of time. The Perks prior, $\alpha_{I,D,k} = \frac{1}{11}$, was chosen as it has little influence over the shape of the posterior. [5]

A number of methods for prioritizing districts were explored. First-responders almost always like to target areas with the heaviest damage first, no matter where they occurred, then visit areas with diminishing levels of destruction. The expected value metric,

$$\mathrm{E}(I_D) = \sum_{k=0}^{10} k d_{I,D,k},\ d_{I,D,k} \sim \mathrm{Beta}(\alpha_{I,D,k}, -\alpha_{I,D,k} + \sum_k \alpha_{I,D,k}), \tag{6}$$

fares poorly as heavily damaged sections within a district are masked by lightly-damaged ones; as an example, Palace Hills during the primary event is not highly prioritized by this metric despite some areas suffering heavy damage. Assigning districts and infrastructure based on the fraction of reports that indicated level-ten damage placed too much emphasis on districts with large uncertainties, as the prior tended to inflate that category. It also poorly described small-uncertainty districts which had an abundance of level-eight and nine reports but few tens. We found the most intuitive damage metric was a power series,

$$\mathrm{Damage}(I_D) = \sum_{k=0}^{10} d_{I,D,k} \cdot s^k \tag{7}$$

with $s = 2$, as this provided a balance between emphasizing reports of heavy damage, filtering out the effects of large uncertainties, and giving some weight to reports of extensive but not heavy damage.

*e-mail: hhornbec@ucalgary.ca
†e-mail: ualim@ucalgary.ca

Nonetheless, we provided users the ability to switch between the expected value metric, a power series with $s = 2$, and a power series with $s = 10$ in lieu of a strict emphasis on the highest category, with the relative repair graphs.

1.2 Layout

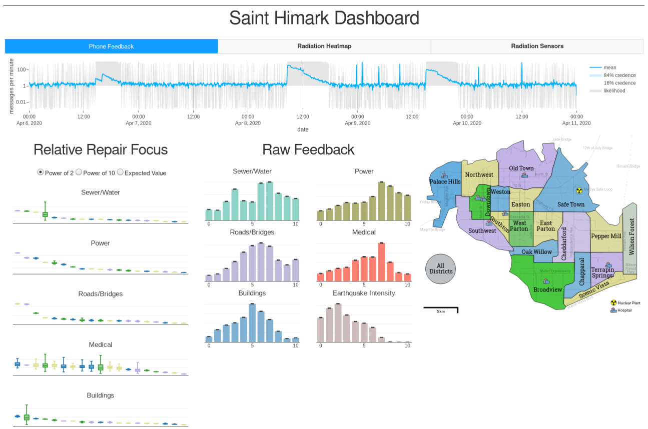

Figure 1: Our full visualization of the dataset for Mini-Challenge 1.

Figure 1 provides a visual overview of our dashboard. As the answer sheet details, districts on the map can be selected by clicking them. The line chart represents the best guess for the current message rate in the selected districts, with light blue bands representing 16/84 credible intervals.[1] The gray area represents the metric captured by Equation 4, the likelihood that the current packet rate estimate is above or below the nominal rate. A subset of time can be selected by clicking and dragging on the desired range, and the full view restored by double-clicking. The Raw Feedback section charts the reports from RUMBLE for the selected districts and over the given timeframe. The Relative Repair Focus section attempts to prioritize the emergency response, again over the selected districts and timeframe.

1.3 Findings

Our answer sheet provides our full findings and many more charts; for this summary, we will excerpt key portions of it.

There were three event clusters in the dataset. The first was a pair of small shocks at approximately 2:30pm on April 6th. Damage was minimal, and primarily concentrated in buildings.

The second event was the primary earthquake, at approximately 8:35am on April 8th. Nearly all infrastructure suffered some level of damage, most notably the power systems. Four spikes in message rate occurring well after the event are due to delayed messages being delivered all at once, and signal the time four districts regained communication. The outage in Terrapin Springs was relatively short, so it was masked by the flood of citizen reports in other districts. The third event was an aftershock of nearly the same magnitude as the original quake, at approximately 3:00pm on April 9th. There's evidence of catastrophic failure within the sewer, power, and road systems. Two post-event message spikes are visible, indicating more communication outages. All these happened hours after the event, however, so the majority of reports had already been sent. Easton seems to have suffered a partial loss, with at least one report getting through during the outage. The outages on the 10th may be due to a failure originating in Safe Town, with little-to-no connection to the earthquake.

[1] These are not to be confused with confidence intervals; see Jaynes *et al.* [3] and Hoekstra *et al.* [2] for further discussion.

While the feedback from most districts is unimodal, there are exceptions such as Cheddarford, Southton, and Safe Town. Palace Hills is the most extreme of these; during the primary earthquake, the feedback is bimodal for all infrastructure types. The divergence is easy to spot in the feedback graphs. While this could be a sign of dirty data, perhaps via a few people entering inaccurate information into RUMBLE, the robust Gaussian behaviour of each peak suggests instead that Palace Hills has heterogeneous infrastructure.

The state of medical infrastructure is reported inconsistently across districts. Districts without a hospital have almost no reporting on the state of their medical facilities. The "About Our City" document only mentions hospitals, so the mostly likely hypothesis is that there are no clinics or other medical infrastructure. This also means medical facility reports from those areas are likely bad data. Rather than filter them out, the wide error bars of both the Raw Feedback and Repair Focus graphs should flag these districts.

1.4 Uncertainty

Our visualization presented uncertainty in a variety of ways. The saturation of each district on the map changes based on the pooled variance of each Dirichlet category for each of the five types of infrastructure, with smaller variances leading to increased saturation. As mentioned, the timeline displays filled bands to represent credible intervals. The Raw Feedback charts include traditional error bars.

The Repair Focus chart details uncertainty through sampling. For the given timeframe and each of the selected districts, the Dirichlet distribution for each type of infrastructure is sampled. The fraction of each damage category is used as input for the chosen damage metric, such as Equation 7, and converted to a single number. Multiple samplings from the Dirichlet create a sample of plausible damage metrics. These are then displayed in a box-and-whisker plot, and sorted so the greatest median is on the left. The larger the box and whiskers, the greater the uncertainty in how damaged that infrastructure is within that district.

2 Conclusion and Future Work

We found our visualization was able to quickly assess the level of damage for each district, and triage repair priorities. Unfortunately, we were unable to assess is the epicentre of any earthquake. While some of our earliest work attempted to answer that, the focus on recovery was judged more important and that component was dropped. The damage metrics were created without feedback from domain experts, and thus their efficiency and accuracy could not be determined. Technical issues caused performance to drop significantly in the deployed app, which could be restored with further effort. Finally, while the underlying algorithms were chosen to be compatible with real-time use, this has not actually been demonstrated.

References

[1] D. Fink. A compendium of conjugate priors. *See http://www. people. cornell. edu/pages/df36/CONJINTRnew% 20TEX. pdf*, p. 46, 1997.

[2] R. Hoekstra, R. D. Morey, J. N. Rouder, and E.-J. Wagenmakers. Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21(5):1157–1164, Oct. 2014. doi: 10.3758/s13423-013-0572-3

[3] E. T. Jaynes and O. Kempthorne. Confidence Intervals vs Bayesian Intervals. In W. L. Harper and C. A. Hooker, eds., *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, The University of Western Ontario Series in Philosophy of Science, pp. 175–257. Springer Netherlands, Dordrecht, 1976. doi: 10.1007/978-94-010-1436-6_6

[4] J. Kerman. Neutral noninformative and informative conjugate beta and gamma prior distributions. *Electronic Journal of Statistics*, 5(0):1450–1470, 2011. doi: 10.1214/11-EJS648

[5] W. Perks. Some observations on inverse probability including a new indifference rule. *Journal of the Institute of Actuaries*, 73(2):285–334, Dec. 1947. doi: 10.1017/S0020268100012270