Predicting Eruptive Events at Volcanoes from Earthquake Data

Ben Mullet **Ankush Singh** **Ying Qi Wong**
Department of Geophysics
Stanford University
{bmullet, ankushs, yqwong}@stanford.edu

1 Abstract

Anticipating the timing and location of volcanic eruptions can reduce the hazard to nearby communities. Many volcanic eruptions are preceded by unrest that includes increased earthquake rates, pronounced ground inflation and elevated gas emissions. However, the implications of these signals on eruption timing are still unclear. In this report, we test a few machine learning algorithms to elucidate the relationship between earthquakes and volcanic eruptions at Kilauea, a highly active volcano in Hawaii. The machine learning algorithms employed include logistic regression, random forests, k-means clustering, and neural networks. First, we use characteristics of individual earthquakes to predict if the volcano is contemporaneously erupting (binary classification). Second, we attempt to predict the time until an eruption begins (regression). The random forest demonstrates some success in eruption forecasting from the earthquake catalog. Future progress in eruption forecasting requires additional datasets or extracted features beyond the historical earthquake catalog.

2 Introduction

Over 600 million people live in close proximity to a volcano and are exposed to the hazards they present [1]. Due to the risk volcanoes pose, accurate and useful forecasts of volcanic eruptions are "a central goal of volcanology" [2]. While most eruptions are proceeded by significant unrest, accurate predictions of volcanic activity, even at very active and well monitored volcanoes, remain elusive.

Kilauea volcano is a large basaltic volcano on the Big Island of Hawaii. Kilauea is well monitored through a network of seismometers, GPS receivers and gas emissions detectors. Kilauea frequently erupts, which poses a threat to residents of the Big Island. In this work, we explore the use of the earthquake record to predict volcanic activity at Kilauea. Physically, volcanic activity can change the stress distribution in the crust which can trigger earthquakes. The models seek to predict Kilauea's eruptive status. For training purposes, historical eruption data was obtained from the compiled Kilauea eruption history as provided by the Hawaii Center for Volcanology.

We investigate two distinct tasks in this research effort. In the first, we use each earthquake's characteristics to predict if Kilauea is erupting at the time of the earthquake ("contemporaneous eruption classification task"). In the second, we perform a forward looking forecast by using the earthquake characteristics to predict the time until an eruption commences (the "time to eruption regression task"). This time is zero for the case that an eruption is already underway.

3 Related work

Despite concerted effort over the last 50 years [2, 3], accurate volcanic eruption forecasts remain elusive. Although significant progress has advanced our understanding of the basic processes that govern eruptions, scientists still struggle to precisely predict the timing of individual eruptions, a point tragically reinforced by the unexpected December 2019 eruption of White Island in New Zealand, which killed over a dozen people despite constant monitoring of volcanic signals [4].

In the rare cases when accurate forecasts have been achieved, success has generally been derived from empirical identification of characteristic patterns in pre-eruptive seismicity and ground deformation. For example, the 2000 eruption of Hekla was accurately predicted to almost the exact minute by Icelandic scientists who observed a pattern of unrest which closely matched the 1991 eruption of the same volcano [5]. More broadly speaking, statistical approaches to volcano eruption forecasting have had limited success [6]. To our knowledge, only one previous study has incorporated modern machine learning techniques to characterize volcanic eruptions. [7] showed that the eruptive state of Piton de la Fournaise could be determined from building a gradient boosted decision tree around the data from a single seismic station.

4 Dataset and Features

The earthquake catalog is from the Advanced National Seismic System (ANSS) and hosted by the World Volcano Database (WOVOdat). Each earthquake has the time, longitude, latitude, depth and magnitude. We filter the earthquakes by location to extract earthquakes along the East Rift Zone, where the eruptions are located at Pu‘u O‘o crater. This gives us 3764 earthquakes in the catalog (Figure 1). From the catalog, we derive the additional feature of earthquake rate, in line with our physical interpretation that elevated volcanic activity should alter the stress in the crust and trigger earthquakes.

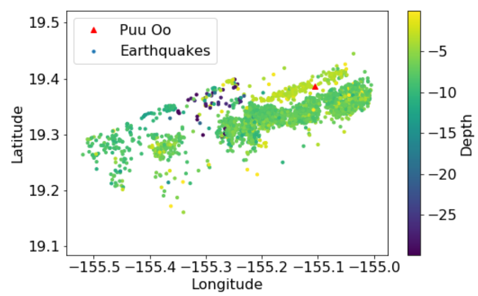

Figure 1: Map of earthquakes along the East Rift Zone of Kilauea. Colors indicate earthquake depths and the red triangle is Pu‘u O‘o.

To summarize, the features employed in this analysis are **longitude, latitude, depth, magnitude, earthquake rate in terms of counts in the last day, last 7 days and last 30 days**. For labels, we either use **erupting (1) or repose (0)** in the classification part or **time to eruption** in the eruption forecasting part. We scale all the features and time to eruption to ensure that they are distributed around mean 0 and standard deviation 1. This is critical for algorithms that employ distances (e.g. K-means) to ensure accurate distance calculations. We separated the data into a training (70% of data), development (20%) and test set (10%) while keeping proportions of eruptions and repose constant.

5 Methods

5.1 Logistic regression

Logistic regression was only used in the contemporaneous eruption classification task using the features listed above.

5.2 K-Means Clustering

For the contemporaneous eruption classification task, we divided the training set into two - one set of earthquakes occurring during repose periods, and one set of earthquakes occurring during eruptive periods. We then performed K-means clustering on each set. We tested different number of classes and found the optimum number of clusters to be 8 (Figure 2).

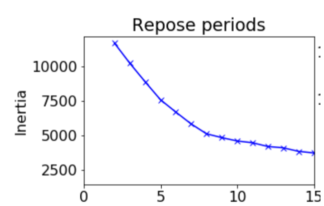

Figure 2: Inertia for different numbers of clusters for K-means clustering of earthquakes during repose periods. There is a clear elbow in the curve at 8 clusters.

For the time to eruption regression task, we perform K-means clustering on the features and time to eruption. The ideal number of clusters in this case is 10. To predict the time to eruption for the development and test sets, we find the closest cluster centroid to the features (i.e. exclude the time to eruption). We assign the time to eruption as the centroid of the nearest cluster.

5.3 Random Forest

We tried the Random Forest for both the classification and forecasting problems as it can take non-linear relationships into account, which reduces bias and also averaging decorrelated trees reduces variance [8]. The number of parameters considered at each split is $\text{int}(\sqrt{p}) = 3$. The hyper parameter tuning was done to find the optimal values of tree depth and number of trees used for the averaging on the basis of the AUROC scores in case of the classification problem (Figure 3)and R^2 values for the regression problem.

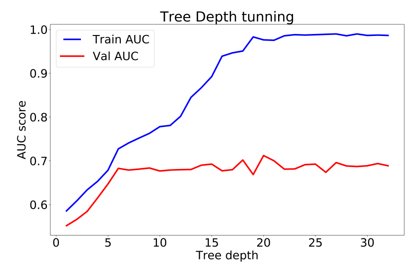

Figure 3: Hyperparameter tuning example for the random forest classification. The maximum tree depth in this case chosen to be 15 on the basis of the development set plateau.

5.4 Neural Network

A simple four hidden layer fully-connected neural network was trained for both tasks. Each hidden layer of the network contained one thousand nodes, such that the model contained roughly four million trainable parameters. Models were built using PyTorch and implemented in the Google compute cloud platform.

For the eruption classification task, we used a final node with sigmoid activation and binary cross entropy loss,

$$\ell_{\text{BCE}} = \frac{1}{N} \sum_i y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})$$

where there are N samples in a batch, and the hat denotes model predictions. For the forecasting task, we used a linear activation function for the final node and mean squared error for loss,

$$\ell_{\text{MSE}} = \frac{1}{N} \sum_i \left(y^{(i)} - \hat{y}^{(i)} \right)^2 .$$

Both models were regularized using weight decay and trained using stochastic gradient descent with batch size of 50. Multiple tests were performed to optimize hyperparameters such as learning rate and weight decay parameter using the validation set.

5.5 Metrics

Our dataset exhibits class imbalance, with approximately 80% of labels being negative (no eruption). We therefore chose to compare the performance of our models using metrics other than classification accuracy. For the contemporaneous eruption classification task, we used two metrics, Cohen's Kappa Coefficient and Area Under the Receiver Operating Characteristic Curve (AUROC). AUROC measures the probability that the model will assign higher a score to a random positive sample compared to a random negative sample. A score of 0.5 represents a meaningless model, whereas a score of 1 is a perfect model. Cohen's Kappa is a modified version of accuracy which takes into account the probability of an accurate test occurring by random chance alone. Thus a Cohen's Kappa score of 0 represents the performance of a model expected by random chance and a score of 1 denotes a perfect model.

For the time to eruption regression task we used root mean squared error and the R^2 value for the plot of observed vs. predicted times.

6 Experiments/Results/Discussion

6.1 Contemporaneous Eruption Classification Task

Comparing all the machine learning methods, the random forest does the best at using earthquakes to classify if the volcano is erupting or not (Figure 4). The random forest classifier has a Cohen's kappa score of 0.42 which can be interpreted as moderate agreement [9]. In comparison, the other methods showed some discriminatory capability at classifying earthquakes as occurring during erupting or not erupting periods, but did not achieve the same level of success on the test set.

The neural network achieved significantly lower performance on this task compared to the random forest. This could be due to a number of factors. Future work should explore alternative choices for loss functions (perhaps better accounting for the imbalanced nature of the dataset), alternative model architectures, and additional exploration of hyperparameters.

Algorithm	Dataset	Kappa	AUROC	Confusion matrix
Logistic Regression	Train	0.37	0.64	[[2165 45] [295 128]]
	Dev	0.25	0.59	[[630 17] [83 22]]
	Test	0.28	0.60	[[315 9] [42 13]]
K-means	Train	0.31	0.67	[[1895 315] [219 204]]
	Dev	0.31	0.67	[[559 88] [54 51]]
	Test	0.24	0.63	[[275 49] [32 23]]
Random Forest	Train	0.84	0.89	[[2043 0] [0 590]]
	Dev	0.46	0.66	[[575 9] [102 66]]
	Test	**0.52**	**0.71**	[[288 5] [50 36]]
Neural Network	Train	0.35	0.63	[[2025 18] [429 161]]
	Dev	0.27	0.60	[[572 12] [131 37]]
	Test	0.29	0.61	[[290 3] [67 19]]

Figure 4: Comparison of different machine learning methods to classify Kilauea earthquakes as occurring during eruption or repose.

6.2 Time to Eruption Regression Task

Again, the random forest does the best to predict the time to eruption (Figure 5). Nevertheless, the correlation between observed and predicted times to eruption remains weak. In the random forest, earthquake rate and latitude are the most important features from this analysis (Figure 6). Regarding the relatively poor performance of the neural network, we again recommend future work should explore alternative choices for loss functions, alternative model architectures, and more exhaustive exploration of hyperparameters.

6.3 Summary

Our results highlight the robust nature of random forest models in achieving high levels of success on relatively distinct tasks (regression and classification). Alternatively, our modified k-means approach seems to work much better for classification than regression, which is perhaps unsurprising given the necessarily discrete nature of the method's predictions. The neural network achieved moderate success on both tasks, but our results indicate future model tuning is necessary for robust performance.

Algorithm	RMSE	R^2
K-means (train/dev/ test)	0.95/0.95/1.02	0.097/0.059/0.037
Random Forest (train/dev/ test)	**0.16/0.68/0.64**	**0.82/0.37/0.38**
Neural Net (train, dev, test)	0.86/0.92/0.90	0.26/0.15/0.19

Figure 5: Comparison of different machine learning methods to forecast the time to eruption based on Kilauea earthquakes.

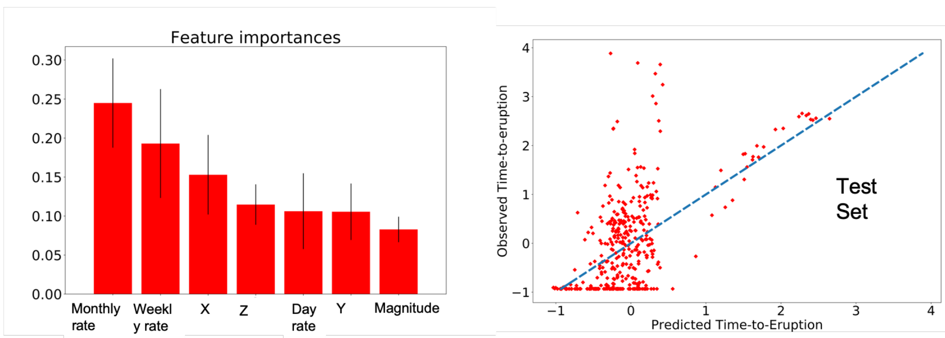

Figure 6: Random forest results with relative feature importance on the left and comparison of the predicted and observed time to eruption values on the right

7 Conclusion/Future Work

From the present analysis, we can conclude that some methods particularly the Random Forest may have predictive capabilities for volcanic eruption. The correlation for time to eruption is weak from all methods and it is clear that additional features will need to be incorporated in addition to interpreted catalog earthquakes. We propose to incorporate additional parameters including earthquake focal mechanisms, low frequency ground deformation data, continuous background geo-phone records and gas emission data. We also plan to expand the dataset to incorporate a larger time period of analysis from 1960-2016 with the caveat that additional complications are introduced by the evolution of recording instruments and processing technology of the time period of analysis. Additionally, including data from tectonically analogous volcanic system from other parts of the world can also be considered to augment the dataset.

8 Code repository

The code for the analysis described in this write-up can be found at the linked repository: `https://github.com/bmullet/PEEVED.git`

9 Contributions

All of us contributed to this research effort. Ankush downloaded the earthquake and eruption data and conducted the random forest. Ben developed the neural network and Ying-Qi performed the K-means clustering analysis.

References

[1] Katharine V Cashman and R Stephen J Sparks. How volcanoes work: A 25 year perspective. *GSA Bulletin*, 125(5-6):664–690, 2013.

[2] R Steve J Sparks. Forecasting volcanic eruptions. *Earth and Planetary Science Letters*, 210(1-2):1–15, 2003.

[3] RW Decker. State-of-the-art in volcano forecasting. *Bulletin Volcanologique*, 37(3):372–393, 1973.

[4] BBC. White island volcano: New zealand troops airlift six bodies off island. *BBC*, Dec 2019.

[5] Ragnar Stefánsson. *Advances in earthquake prediction: research and risk mitigation*. Springer Science & Business Media, 2011.

[6] Andrea Bevilacqua, Raffaele Azzaro, Stefano Branca, Salvatore D'Amico, Emanuela De Beni, Franco Flandoli, and Augusto Neri. Estimates and models of the statistical dependence between local earthquakes and flank eruptions at mt. etna volcano (italy): an old topic revised through new historical data. In *AGU Fall Meeting 2019*. AGU, 2019.

[7] Christopher X Ren, Aline Peltier, Valerie Ferrazzini, Bertrand Rouet-Leduc, PA Johnson, and Florent Brenguier. Machine learning reveals the seismic signature of eruptive behavior at piton de la fournaise volcano. *arXiv preprint arXiv:1909.12395*, 2019.

[8] Daphne Koller, Nir Friedman, Sašo Džeroski, Charles Sutton, Andrew McCallum, Avi Pfeffer, Pieter Abbeel, Ming-Fai Wong, David Heckerman, Chris Meek, et al. *Introduction to statistical relational learning*. MIT press, 2007.

[9] Nicole J-M Blackman and John J Koval. Interval estimation for cohen's kappa as a measure of agreement. *Statistics in medicine*, 19(5):723–741, 2000.