Use of Machine Leaning to Analyze and – Hopefully – Predict Volcano Activities

Justin Parra[1], Olac Fuentes[1], Elizabeth Anthony[2],
and Vladik Kreinovich[1]
[1]Department of Computer Science
[2]Department of Geological Sciences
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
jrparra2@miners.utep.edu,
ofuentes@utep.edu,
eanthony@utep.edu
vladik@utep.edu

Abstract

Volcanic eruptions cause significant loss of lives and property around the world each year. Their importance is highlighted by the sheer number of volcanoes for which eruptive activity is probable. These volcanoes are classified as in a state of unrest. The Global Volcano Project maintained by the Smithsonian Institution estimates that approximately 600 volcanoes, many proximal to major urban areas, are currently in this state of unrest. A spectrum of phenomena serve as precursors to eruption, including ground deformation, emission of gases, and seismic activity. The precursors are caused by magma upwelling from the Moho to the shallow (2-5 km) subsurface and magma movement in the volcano conduit immediately preceding eruption.

Precursors have in common the fundamental petrologic processes of melt generation in the lithosphere and subsequent magma differentiation. Our ultimate objective is to apply state-of-the-art machine learning techniques to volcano eruption forecasting. In this paper, we applied machine learning techniques to the precursor data, such as the 1999 eruption of Redoubt volcano, Alaska, for which a comprehensive record of precursor activity exists as USGS public domain files and global data bases, such as the Smithsonian Institution Global Volcanology Project and Aerocom (which is part of the HEMCO data base). As a result, we get geophysically meaningful results.

1 Volcano Eruption Forecasting: Formulation of the Problem and State-of-the-Art

Need for volcano eruption forecasting. Because of the possible catastrophic consequences, researchers have always been trying to develop methods for predicting volcano eruptions.

We believe that volcano eruption forecasting is possible. The hope for predicting volcano eruptions comes from the fact that most eruptions are preceded by different types of unusual activities.

In geophysical terms, volcanoes that will erupt in the near future are classified by the community of volcanologists as in a state of "unrest". Unrest is manifested as a combination of changes in the amount and chemical composition of volcanic gas emissions [13], ground deformation above the volcanic edifice [2], and seismic activity [10]. The activity is the result of subsurface movement of magma as it ascends to the surface.

Volcano eruption forecasting is difficult. Unfortunately, in spite of the seemingly clear relation between these precursors and the following eruptions, there is still no good way to make long-term predictions of volcanic activity: no matter what combination of precursors we select:

- sometimes, a similar combination results in an eruption, while
- in other cases, a seemingly similar activity is not followed by an eruption.

Need for probabilistic forecasting. In general, the relation between the precursors and the eruptions has a probabilistic character: the presence of precursors does not necessarily indicate that the eruption is imminent, but it seems to increase the probability of the eruption.

From this viewpoint, we can only predict probabilities of eruptions of different strength and type.

Probabilistic methods of volcano eruption forecasting: state-of-the-art. Several research papers use probabilistic methods to predict the eruption probabilities; see, e.g., [1, 9, 12].

These methods start with the known power-law models that describe the relation between the different characteristics – e.g., between the eruption strength and the time to the next eruption – and add appropriate probabilistic models to describe the inaccuracy of these relations. The parameters of the corresponding multi-parameter models are then tuned to match the observed phenomena. The resulting tuned model is then used for forecasting.

This statistical approach works perfectly well in many applications to engineering and science. For volcanic eruptions, this approach has led to several reasonable short-term and long-term probabilistic forecasts.

However, these predictions are still far from perfect. It is therefore desirable to improve the accuracy and reliability of the existing predictions.

2 Analysis of the Problem and the Resulting Ideas

Why predicting volcano eruptions is different from other types of predictions. In our opinion, two specific features of volcanic eruptions limits the potential of such purely statistical approach.

First specific feature of predicting volcano eruptions. The first specific feature of volcano eruptions is related to the fact that successful statistical methods require that we know the parameters of the corresponding probabilistic models.

To accurately determine the values of these parameters in a statistical setting, we need to have reasonably large data samples. This is a big problem for volcanic studies, since, in contrast to many engineering and scientific phenomena, volcano eruptions are relatively rare events.

What has been done to overcome this difficulty: Bayesian approach. One approach to compensate for the smallness of samples is to add expert knowledge, which can be described in terms of subjective prior probabilities of different events.

These approximate prior values of the corresponding probabilities are then updated based on the observations; the formulas for such an update were first discovered by Bayes; because of this fact, such an approach is known as *Bayesian*; see, e.g., [12].

Limitations of the Bayesian approach. The problem with applying Bayesian approach to volcanic eruptions is that different experts may have different opinions, so we end up with different prior probabilities – and thus, different predictions.

When we have a reasonable large data sample, the observation-based update tilts the original subjective probabilities towards the observed frequencies. As a result, the dependence on the initial (prior) probabilities drastically decreases.

However, for situations like volcanic eruptions, when the sample sizes are small, the resulting predictions remain strongly dependent on the original subjective probabilities.

Second specific feature of predicting volcano eruptions. The second specific feature of predicting volcano eruptions is as follows.

In many engineering and scientific phenomena, when we know reasonably accurate formulas describing the dependence between different quantities – e.g., differential equations describing elasticity, Navier-Stokes equations describing liquids, etc.

In contrast, for volcanic activities, we do not know the exact shape of the corresponding dependencies.

When we use the traditional finite-parametric probabilistic models, e.g., power law models (which are known to be a rather crude approximation to real-life phenomena), we are thus limiting ourselves to these crude models, and hence, restricting our ability to forecast.

How to overcome the corresponding difficulty: need for machine learning techniques. To overcome this problem, it is therefore desirable to use non-parametric prediction models.

Such methods, when we do not fix the shape of the dependence from the very beginning, but let the data determine this shape, are known as *machine learning* techniques; see, e.g., [3].

When we apply such techniques, then, instead of a researcher trying to guess the corresponding relation – such as a power law – the computer-based system determines this relation by itself, based only on the observations.

Machine learning algorithms start the observed data: both

- the values of the quantities that we want to predict and
- the values of the possible related quantities that we would like to use in this prediction.

Based on this data, machine learning algorithms eventually come up with a computer model that makes accurate predictions in all given situations – and, in many applications, makes successful predictions in new situations as well.

Machine learning techniques are currently ubiquitous in many applications, they underlie the ability of modern cellphones to recognize voices, they provide security against hackers and spam, they are behind the recent successes of Artificial Intelligence such as computers winning over Go masters, and many other applications; see, e.g., [3, 5, 11].

Our eventual goal is to apply machine learning techniques to the volcanic data to come up with effective forecasting techniques.

Why we believe that machine leaning methods will be helpful in volcano eruption forecasting. Our belief in machine learning techniques comes not only from their successes in modern appliances, but also from our experience of successful using them in different applications.

In our previous research efforts, we have used these techniques to predict the best strategy for a robot [7, 8], to determine the parameters of stellar atmospheres based on astronomic observations [6], and in many other applications.

3 Our Study: Description and Results

Description of the problem. To test our belief, we did some preliminary proof-of-concept analysis.

Specifically, for two volcanoes for which there is an extensive record of small nearby earthquakes – Popo in Mexico and and Readout in Alaska – we analyzed the spatial locations of these earthquakes in comparison with the location of the volcano itself.

What data we used. In this study, we use open source data of precursor activity for the Aleutian chain of volcanoes [4].

The Aleutians are an arcuate chain of active volcanoes that reaches from Alaska to Russia. They represent the subduction (underthrusting) of Pacific

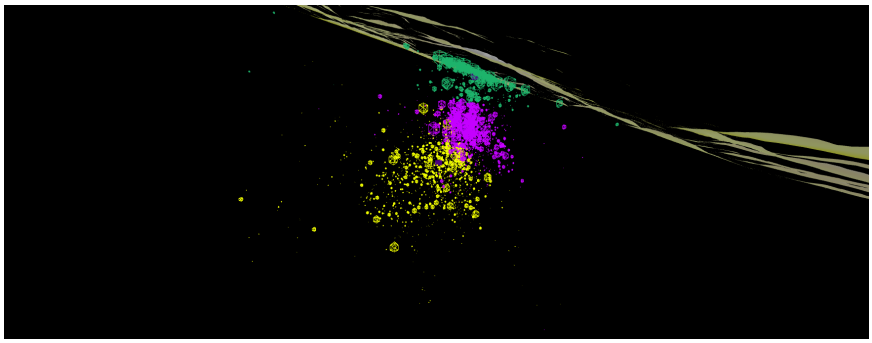

Figure 1: Clusters of earthquake locations

lithosphere beneath North America. Because of their location, silicate ash erupted from them into the atmosphere impacts air traffic across major flight paths in the Pacific. We have begun our analysis with the seismic record for the volcanoes to maximize the data elements available.

What data processing methods we used. We started with the simplest type of learning, when instead of trying to predict the numerical value of a real-valued quantity, we try to predict a simple quantity with a very small number of possible values.

In such a prediction, we thus classify different objects or events into one of the few groups – corresponding to different values of the predicted few-valued quantity.

In other words, we cluster the events or objects into a small number of clusters, so that ideally,

- the events/objects within each cluster are similar to each other, while
- events/objects from different clusters are different.

For this pilot study, we use one of the simplest clustering algorithms – k-means.

First result and its geophysical interpretation. For each volcano, when we applied this clustering algorithm to the locations of all the nearby earthquakes, these locations formed three clearly distinguished clusters.

Earthquakes from the first two clusters are mostly vertically located right beneath the volcano. Depth-wide, they seem to correspond to the volcano pipe and to the place where the magma goes from the magma chamber into the pipe.

The third, deeper cluster is spread mostly horizontally, it seems to correspond to a sill-shaped magma chamber.

Interestingly, the center of this third cluster is shifted in comparison to the volcano itself, so that the volcano is approximately at the edge of the cluster. In other words, it looks like the magma accumulated in the magma chamber finds

the way up along the edges of the chamber – which seems to be in good accordance with the observed asymmetry of volcanic eruptions, which also usually start not at the center of the volcano, but on one of the edges of the volcano's throat (which explain the visible asymmetry of many volcanic calderas).

Second result and its geophysical interpretation. When we repeated our clustering experiment, this time only considered earthquakes above certain level, then:

- we still got three clusters at different depths, but
- this time, all three clusters were vertically aligned.

This also makes geophysical sense:

- while we have seismic activity throughout the whole magma chamber,
- we expect stronger activity in locations were the magma is most active, i.e., in the location where the magma is going up – which is directly beneath the pipe.

Conclusion. Of course, these are preliminary results that need to be further analyzed and confirmed.

However, the very fact that, without inputting any geophysical knowledge into our computations, by simply applying general algorithms to observed data, we got geophysically meaningful results, makes us confident that by applying more sophisticated machine learning techniques to volcanic data, we will be able to capture the corresponding geophysical phenomena and thus, make reasonable forecasts.

Acknowlegments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

References

[1] M. S. Bebbington and W. Marzocchi, "Stochastic models for earthquake triggering of volcanic eruptions", *Journal of Geophysical Research*, 2011, Vol. 116, Paper B05204.

[2] J. Biggs, E. Y. Anthony, and C. J. Ebinger, "Multiple inflation and deflation events at Kenyan volcanoes, East African Rift", *Geology*, 2009, Vol. 37, pp. 979–982.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[4] K. F. Bull and H. Buurman, "An overview of the 2009 eruption of Redoubt Volcano, Alaska", *Journal of Volcanology and Geothermal Research*, 2013, Vol. 259, pp. 2–15.

[5] K. Buza and J. Koller, "Classification of Electroencephalograph Data: A Hubness-Aware Approach", *Acta Polytechnica Hungarica*, 2016, Vol. 13, No. 2, pp. 27–46.

[6] O. Fuentes, "Automatic Determination of Stellar Atmospheric Parameters Using Neural Networks and Instance-Based Learning", *Experimental Astronomy*, 2001, Vol. 12, No. 1, pp. 21–31.

[7] O. Fuentes and R. C. Nelson, "Learning Dextrous Manipulation Strategies for Multifingered Robot Hands Using the Evolution Strategy", *Machine Learning*, 1998, Vol. 31, pp. 223–237.

[8] O. Fuentes and R. C. Nelson, "Learning Dextrous Manipulation Strategies for Multifingered Robot Hands Using the Evolution Strategy, *Autonomous Robots*, 1998, Vol. 5, pp. 395–405.

[9] W. Marzocchi and M. S. Bebbington, "Probabilistic eruption forecasting at short and long time scales", *Bulltin of Volcanology*, 2012, Vol. 74, pp. 1777–1805.

[10] J. A. Power, S. D. Stihler, B. A. Chouet, M. M. Haney, and D. M. Ketner, "Seismic observations of Redoubt Volcano, Alaska – 1989-2010 and a conceptual model of the Redoubt magmatic system", *Journal of Volcanology and Geothermal Research*, 2013, Vol. 259, p. 14.

[11] S. Preitl, R. E. Precup, and Z. Preitl, "Development of conventional and fuzzy controllers and Takagi-Sugeno fuzzy models dedicated for control of low order", *Act Polytechnica Hingarica*, 2005, Vol. 2, No. 1, pp. 75–92.

[12] R. Tonini, L. Sandri, D. Rouwet, C. Caudron, W. Marzocchi1, and Suparjan, "A new Bayesian Event Tree tool to track and quantify volcanic unrest and its application to Kawah Ijen volcano", *Geochemistry, Geophysics, Geosystems*, 2016, Vol. 17, pp. 2539–2555.

[13] C. Werner, P. J. Kelly, M. Doukas, T. Lopez, M. Pfeffer, R. McGimsey, and C. Neal, "Degassing of CO_2, SO_2, and H_2S associated with the 2009 eruption of Redoubt Volcano, Alaska", *Journal of Volcanology and Geothermal Research*, 2013, Vol. 259, pp. 270–284.