Volcanic Eruption Prediction

Matt Barrett | https://github.com/mattb1888/cs5665_project

Problem

The goal of this challenge was to predict how long it be until a volcano erupted. Our submissions were evaluated on the mean absolute error between our predictions and the actual value

Data

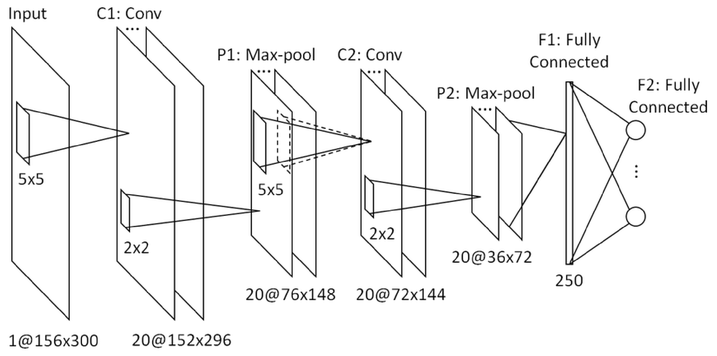
The data given was 10 minutes of readings taken every 1/100th of a second from 10 sensors around the volcano, for a total of 60,000 data points per volcano. We were provided with training data from 4,300 volcano including the time until the volcano erupted. We were also given 4,500 volcanoes for which we were to predict the time until eruption.

Approach

I took several approaches to the problem. I tried extracting features from the data two different ways and I used both neural networks and random forests on each set of extracted features. The most successful feature extraction was to get the mean, standard deviation, and every 10th percentile from 0 to 100, from each sensor for a total of 130 features per volcano.

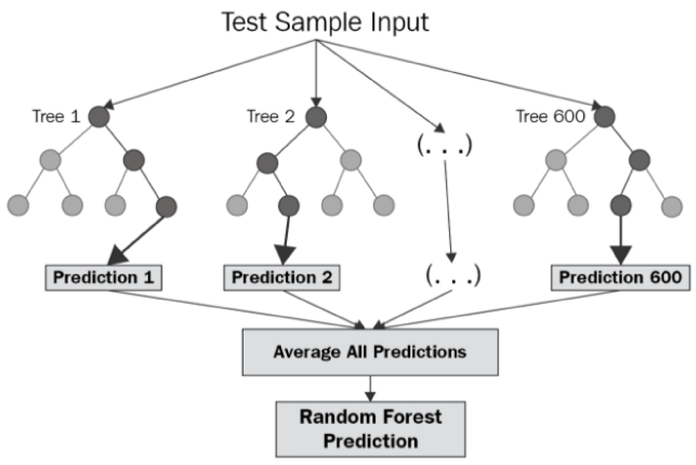
CNNs

I tried a variety of fully connected and convolutional neural networks. The best results came from a neural network made up of two pairs of convolutional/pooling layers followed by four fully connected layers



Random Forests

I tried random forests made up of increasing numbers of decision trees until I found where accuracy leveled off as random forests tend to do. In this case maximal accuracy was reached at around 250 decision trees.



Results

My random forests significantly out performed my neural networks. In my testing the best random forest had an MAE of 5.3 million, while my best neural network had an MAE of 10.6 million. The random forests also are considerably quicker to train and predict.

Conclusions

I believe that random forests work better on this problem than neural networks is because I am already extracting all the features from the data. This is where random forests excel. Neural networks are best when you do not know which features are important, e.g. image classification. My current best score on the Kaggle competition is 264 out of 358 with an MAE of 7.7 million

264	**Matthew Barrett**	7755572

Source Code

https://github.com/mattb1888/cs5665_project

I am also still updating this poster so you can check there for the latest version

Image Citations

https://miro.medium.com/max/700/1*ZFuMI_HrI3jt2Wlay73IUQ.png

https://www.researchgate.net/profile/Shun_Miao/publication/280538045/figure/fig2/AS:354647291252736@1461565911272/Structure-of-the-multi-task-learning-convolutional-neural-network.png