Gabrielle Merritt
Elizabeth Reeves
Levon Avakian
CIS 519

California Wildfire Prediction Using Machine Learning Techniques

Our project aims to create a learning algorithm that can predict the likelihood, location, and severity of wildfires in the state of California. On average, the state of California loses over 100 million dollars and 218,000 acres[1] due to wildfire damages. Currently California only employs fire prevention methods such as restricting certain kinds of fuels, controlled fires, and fire education to curb the damage of wildfires; however, using machine learning to learn fire patterns can help fire departments properly allocate resources and take targeted measures to preventing large wildfires.

In our initial research, we came across several papers describing modeling and machine learning techniques to predict wildfires. Most of the predictive methods we found dealt exclusively with weather forecasting models, weather based learners[2], or human factors based learners[3]. Currently, only the forecasting models have been used in practice (such as those employed by the National Interagency Fire Center to create qualitative seasonal outlook reports[4]), while machine learning techniques have mainly been used to verify the accuracy of the forecasting models with poor to moderate success[3]. We have yet to encounter a system that combines human factor features with weather based features. The methods we have encountered use a very small feature set compared to their instance data, and as a result the majority of the analysis was done using support vector machines or logistic regression.

We plan on using weather data from the National Weather Service as well as fire data from the California Department of Forestry and Fire Protection, as both of these sources have data dating back several decades. For our approach our feature set will be much larger compared to the previously discussed techniques, but compared to our instance set, the feature space will be orders of magnitude smaller. Therefore, for our base learner we are considering using support vector machines with either Gaussian kernels or no kernels, logistic regression, or decision trees. Another machine learning algorithm we are considering is k-means. Since we want to show a heat map of large wildfire probability as our outcome, we would use k-means as a way to divide the data into clusters with similar features (such as similar geographical areas or adjacent counties), and then use a secondary machine learning algorithm to learn and predict over the clusters.

A positive outcome from this project could benefit a number of organizations, most notably the California Department of Forestry and Fire Protection and other members of the California Fire Alliance, which is devoted to preventing and fighting wildfires in the state of California. The results of our algorithm could also indirectly benefit campers and residents of California by reducing fire damage and costs. Finally, increased control of wildfires could have environmental benefits on a global scale[5].

---

[1] http://www.fire.ca.gov/communications/downloads/fact_sheets/firestats.pdf
[2] http://journals.ametsoc.org/doi/full/10.1175/JAM2513.1
[3] http://www.sciencedirect.com/science/article/pii/S1364815214000814
[4] http://www.predictiveservices.nifc.gov/outlooks/outlooks.htm
[5] http://www.scientificamerican.com/article/fires-fuel-climate-change/