AGNI: Autonomous Geospatial system for Noticing Ignition

Jack Gallagher and Baeo Maltinsky

Problem Statement

Wildfires cause immense economic damage and loss of life across the United States. California is particularly vulnerable due to its fire-dependent ecosystems and large annual variation in rainfall. While naturally occurring wildfires are a normal and necessary part of Californian ecology, they can pose a danger to settlements if they occur near inhabited areas and are allowed to grow out of control. Anthropogenic fires are often even worse, as they tend to begin within or near cities and other critical infrastructure. Heavy winds and other severe weather increase the failure rate of relevant infrastructure such as transmission lines while also intensifying and spreading the fire. We may have seen this in the 2017 Sonoma County Fires[1], and similar incidents have occurred in the past[2].

Early detection and forecasting methods can improve calibration of response to seemingly small incidents, allowing rapid, targeted mobilization of firefighting resources.

Existing Solutions

Fire departments in most of the United States rely largely on models built in the 1960s such as the National Fire Danger Rating System. These models are low resolution, low specificity, and don't take advantage of the quantity of data available to a modern project. While academic models have done better, they are almost always tightly coupled to particular small geographic areas due to data poverty and fragmentation. To quote a review paper on the subject[3]:

> Predictive models have exploited several sources of data describing fire phenomena. Experimental data are scarce; observational data are dominated by statistics compiled by government fire management agencies, primarily for administrative purposes and increasingly from remote sensing observations. Fires are rare events at many scales. The data describing fire phenomena can be zero-heavy and non-stationary

> over both space and time. Users of fire modeling methodologies are mainly fire management agencies often working under great time constraints, thus, complex models have to be efficiently estimated.

What can be done?

We propose a three-pronged approach for dealing with the problems outlined above.

Data Aggregation

Forest fire data is extremely fragmented[4]:

> To conduct even the most rudimentary interagency analyses ... one must harvest records from dozens of disparate databases with inconsistent information content.

Other authors have described the data as "inconsistent and difficult to use"[5]. We would begin by compiling a single, more consistent dataset, containing information about historical location and severity of fires alongside weather, air quality, and climate data. This would allow dramatically improved experimentation and generalization of statistical tools, with the larger dataset providing enough information for more sophisticated models.

Modern Machine Learning Techniques

Once we've compiled a data set, we can begin creating and testing statistical models of fire danger. Simple approaches have been able to get surprisingly good results in small areas with unified data sets. One Turkish study was able to get 83% accuracy using a fully-connected single-layer neural network with only 13 nodes[6], and this seemed broadly representative of both performance and sophistication across the literature[7][8]. We expect that, given a larger and more complete data set, we would be able to replicate and improve on these results across the state, allowing much more detailed risk analysis than is currently possible. Some possible approaches are:

- Spatio-temporal convolutional neural networks
- Recurrent neural networks with convolutional hidden layers
- Generalized additive models
- Autoregressive integrated moving average
- Random decision forests

With proper calibration, we expect any of these models would constitute an improvement over existing systems, though the most suitable approach would have to be found through experimentation.

Existing Sensor Networks

Government agencies such as the EPA[9] and NOAA[10] maintain networks of sensors monitoring environmental conditions such as air quality and metereological data. These networks provide a wealth of data that could be used in the real-time predictive models described above.

Conclusion

Very promising preliminary work has been done building statistical models in this area. There is a near-universal trend of researchers in the literature citing lack of data as a prominent constraint in generalizing their models. Furthermore, the statistical tools applied thus far have been relatively simple, and far behind the state-of-the-art in machine learning. We expect that addressing the issues of data and systematic evaluation of existing approaches would produce a system capable of vastly improved forecasting and calibration of emergency response.

Notes

[1]https://www.mercurynews.com/2018/06/08/pge-blamed-for-multiple-north-bay-wildfires

[2]http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M196/K872/196872312.PDF

[3]http://doi.org/10.1214/13-STS451

[4]https://doi.org/10.5194/essd-6-1-2014

[5]https://doi.org/10.1111/j.1467-9671.2008.01117.x

[6]https://doi.org/10.1080/19475705.2015.1084541

[7]https://doi.org/10.1111/j.1467-9671.2008.01117.x

[8]http://doi.org/10.1214/13-STS451

[9]https://www.epa.gov/aqs/

[10]https://www.ncdc.noaa.gov/data-access