Andrea Cicero

Machine Learning Applied to Geyser Eruptions in Yellowstone 1970-2019

Old Faithful Geyser in Yellowstone National Park is a unique hydrothermal feature that gained its moniker for erupting very regularly every hour. In recent years, the waiting time has fluctuated considerably. Several studies have addressed this relationship using a small dataset of 272 measurements taken in August 1981 (Hardle 1991, Azzalini and Bowman 1990) and is even stored as a stock dataset used in the R program manual. A more recent study examined data from 2000-2011 (Hartigan 2013). This project proposes to compile (1) all eruption data available from ranger log transcriptions from 1970 to July 2019 to create the most complete dataset of Old Faithful eruptions, and (2) apply machine learning techniques to refine the relationship between waiting times and duration and frequency of eruption.

A dashboard is proposed with some background geological information, plots of historical data, application of machine learning techniques (combination of K-means, linear, multivariate normal, SVM or some other model to fit the time/frequency data – may need some help here). Ideally, I would like to pull in live data to update the models and compare against the national park and other group predictions. I would also like to create an API of curated data for others to use. Time permitting, I would also like to examine the relationships between weather/streamflow and/or temperature and how they influence the eruption frequency.

Challenges: ranger log transcriptions are made by multiple individuals and are not continuously monitored (i.e., winter months when the park is closed; in early years nightly observations were not made, a span of years in the 1990s when the great fire interrupted Old Faithful observations).

Resources:

Hardle, W. (1991) Smoothing Techniques with Implementation in S.

New York: Springer.

Azzalini, A. and Bowman, A. W. (1990). A look at some data on the

Old Faithful geyser. Applied Statistics 39, 357-365.

Hartigan J. A. (2013) Old Faithful. Yale University statistics project.

<https://geysertimes.org/>

<http://www.geyserstudy.org/geyser.aspx?pGeyserNo=OLDFAITHFUL>

<https://medium.com/@kasumisanchika/k-means-clustering-for-old-faithful-geyser-eruptions-analysis-74f3df8979b1>