Probabilistic Dust Storm Prediction

Presentation 2 – Methodology

Sean Flaherty

2017

Overview

- Goal: Use data mining/machine learning to create a predictive model for dust events on a local scale.
- Previous research: Univariate predictor (500mB geopotential height) using image processing (ZNCC). Only good at predicting large events [1].
- Justification: Want an accurate predictor for meso- and micro-scale dust events.

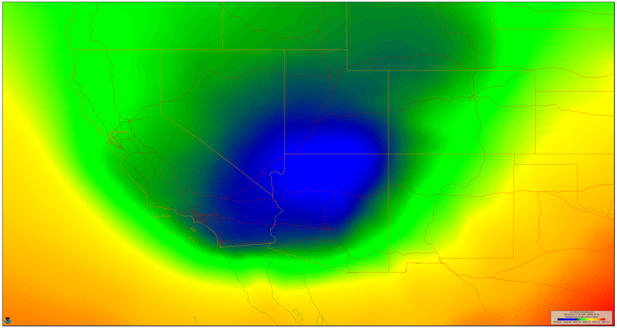
Gathering Data

- RAP/RUC forecast model – NOMADS (NOAA Operational Model Archive and Distribution System).
- Forecasts available from NOAA online repository (HTTPS/FTP).
- Data downloaded using GNU wget utility [2].

Data Format

- GRIB – GRIdded Binary (WMO standard for weather data).
- RAP/RUC models update forecasts hourly with 13 and 25.2 km resolutions. Each file contains forecast models for a single time.
- Each file has a number of weather parameters, each of which with a grid of data points corresponding to locations.

GRIB Example



A single parameter's raster shown using NOAA Weather and Climate Tool. This image shows the 500mB geopotential height at 18:00GMT preceding a dust event on April 14, 2012 [3].

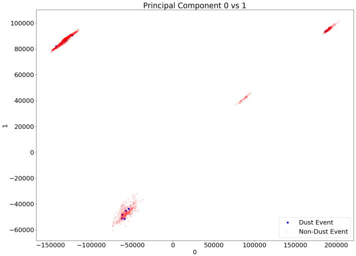
Reading GRIB files

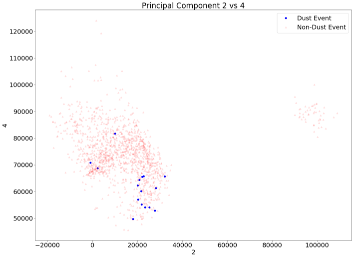
- pygrib Python library – allows opening of .grb, .grb2 files in Python
- Opening a GRIB creates a file iterator, with each object in it a weather parameter [4].
- Each weather parameter has various attributes, including a raster of latitudes/longitudes and data for the parameter.
- Weather data gets stored into CSV files for easier lookup.

PCA

- Principal component analysis reduces the dimensionality of the data.
- Transforms data into subspaces with the most spread between points - explains most of the variance in the data.
- RUC dataset has 315 dimensions for each instance – could make algorithms less effective (curse of dimensionality).

PCA plots



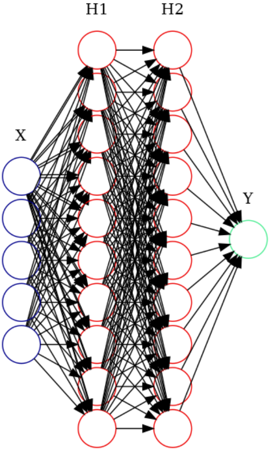


Plots of principal components against each other. Left shows little no distinction between dust and non-dust events, while right shows some.

Algorithms

- Feedforward NN
- Feedforward NN with PCA
- RNN/LSTM
- RNN/LSTM with PCA

Try each algorithm and see which one provides the best accuracy.



Implementation of algorithms

- TensorFlow Python library
- Machine learning utility for creating computation graphs, doing lazy evaluations, using GPU for faster processing.
- Automates backpropagation and optimization algorithms [5].

References

[1] Armenta, Rebecca B. "Geopotential height patterns at 500mb associated with dust storms in the United States/Mexico border region during January-May of 2011-2014." May 2016 New Mexico State University. Access May 31 2017.
[2] "GNU Wget 1.18 Manual." GNU Project. Web. Jun. 27 2017.
[3] "Rapid Refresh (RAP)." National Centers for Envrionmental Information. NOAA. Web. Jun. 27. 2017.
[4] "pygrib documentation." Github. Dec. 29 2014. Web. Jun. 14 2017.
[5] "Getting Started with TensorFlow." TensorFlow. Web. Jul. 5 2017.