August 1st, 2017

Udacity Machine Learning Nanodegree

Igor Oliveira

Capstone Proposal

Machine Learning for Fog Forecasting

Domain Background

This project will attempt to use meteorological observations data to forecast the occurrence of fog, an extremely important low visibility event that have profound impact in air transport operations. This application of Machine Learning has been used before, specially the use of Artificial Neural Networks (ANN) (Refs: [1]-[5]). The existing research in this field provides a good guidance about the type of data to be considering in the training process. It also provides background about the good performance of ANN methods when compared to traditional techniques (Ref: [5]).

As a meteorology researcher, this seemed like a good choice for capstone project in the Machine Learning Nanodegree Course.

Problem Statement

The problem to be solved is the forecast of visibility values (in meters) with a few hours of lead time (from 3 to 12). Airports and other sites report the visibility values, as well as several other atmospheric variables like wind, temperature, humidity and pressure, allowing for the application of machine learning techniques. Since we have the reference use of successful ANN techniques in the prediction of this phenomena, this project will try to reproduce this method and try Deep Neural architectures in search of possible better performance. The results of the regression model proposed for visibility forecast can be verified via metrics like R_squared, Mean Absolute Error or Mean Squared Error. Since the data is available in many sites worldwide, replication is possible. However, since the strong dependence with local information it is possible that model development could vary depending on the site where forecast is being performed.

Datasets and Inputs

There is a public dataset of weather observations provided by several aerodromes (METAR) and other reporting stations (SYNOP) which consist in coded data, captured manually or automatically and containing observed values for several meteorological characteristics such as air temperature, relative humidity, wind speed and direction, atmospheric pressure, and many more. These reports are gathered at hourly time plus at significant weather events (METAR) or 3-hourly (SYNOP)

These datasets can extent for many decades, depending on the history of the aerodrome/station. Many sources can be found online such as the Iowa State University and weather.cod.edu, sometimes already in decoded format. These datasets contain the necessary features necessary for the training, validation and testing according to the references, the following will be used :

- Air Temperature
- Dewpoint Temperature
- Atmospheric Pressure
- Wind Speed & Direction

Three airports were selected initially to perform the training, given the frequent occurrence of fog events in their historical datasets

SBPA - Porto Alegre, Brazil: https://uk.flightaware.com/resources/airport/SBPA/weather

SUMU - Montivideo, Uruguai: https://uk.flightaware.com/resources/airport/SUMU/weather

SAEZ - Buenos Aires, Argentina: https://uk.flightaware.com/resources/airport/SAEZ/weather

Regression techniques will be used by using ML to predict the value of visibility or ceiling reported as targets for the algorithms. It is important to select training, testing and validation sets that include a balanced number of low visibility events to avoid potential biases.

Solution Statement

The solution to the fog forecasting problem will be achieved through regression. Difference architectures of Neural Networks will be tested, using the value of visibility as targets for training and testing, since this is the typical architecture provided in the references. There is a specific opportunity here to pioneer the use of deep neural networks for fog forecasting, as not many references with this approach were found.

Benchmark Model

The references cited in this proposal (Refs [1] to [5]) provide different benchmarks results for fog forecasting in different aerodrome. Costa et al (2006), for example, provides reasonable values of Root Mean Square Error (RMSE) for prediction of visibility values. These works use the same type of data considered in this proposal and allow for a reasonable comparison.

Evaluation Metrics

The main evaluation metric to be considered is the Root Mean Squared Error, RMSE:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N} \left| \frac{a_k - y_k}{a_k} \right|^2}{N}}$$

Where,

a_k = Wanted exit or observation

y_k = Calculated exit

N = Number of forecasts

Project Design

An initial step necessary in this work is to analyze aerodromes amongst the total available globally and find some with a reasonable number of fog observations, in order to derive suitable training sets. The sites where rare fog or low visibility events are observed might not provide a consistent dataset. This will result in the selection of a few sites for forecasting.

After that, we should organize the data in order to match the fog observations and the visibility values with the corresponding meteorological features. After some data cleaning, which includes checks for missing and unreasonable data points, this will generate the proper dataset for training, validation and testing. Next, regression starts properly with the application of different Neural Networks architectures including DNNs for prediction of visibility values. The performance of the machine-learning algorithms applied will be measured according to the indicated verification metric.

References:

1. Pasini, Antonello, Vinicio Pelino, and Sergio Potestà. "A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables." Journal of Geophysical Research: Atmospheres 106.D14 (2001): 14951-14959.

2. Costa, Saulo B., et al. "Fog forecast for the international airport of Maceió, Brazil using artificial neural network." Proc. 8th ICSHMO, Foz do Iguac u, Brazil (2006): 24-28.

3. Gultepe, I., et al. "Fog Research: A Review of Past Achievements and Future Perspectives." Pure and Applied Geophysics 6.164 (2007): 1121-1159.

4. Bremnes, John Bjørnar, and Silas Chr Michaelides. "Probabilistic visibility forecasting using neural networks." Pure and Applied Geophysics 164.6-7 (2007): 1365-1381.

5. Marzban, Caren, Stephen Leyton, and Brad Colman. "Ceiling and visibility forecasts via neural networks." Weather and forecasting 22.3 (2007): 466-479.