Prospectus

Deep Learning techniques applied to automatic flagging of navigational, meteorological and oceanographical observations from NOAA research vessels.

1. Abstract:
   1. Discuss the research question:
      1. The Shipboard Automated Meteorological and Oceanographic System (SAMOS) initiative improves the quality of navigational, meteorological and oceanographic observations collected on research vessels through the use of a number of quality control procedures. One such procedure is a visual quality control method which involves individual analysts manually reviewing large amounts of raw observational data. This method of quality control requires a considerable amount of time on the part of the analysts. Visual quality control is nonetheless a vital part of the quality control process as traditional automated quality evaluation methods are relatively limited in their ability to evaluate data holistically. The labor involved in this process and the potential to develop other non-traditional methods for evaluating data has inspired this research project which aims to develop a neural capable of providing anomaly detection the data with comparable (if not superior) precision than our human analysts.
   2. The Data
      1. The dataset consists of “observation data”, collected via meteorological instruments installed on research vessels. Since the project’s launch in 2005, 52 research vessels have contributed to the Samos dataset. There are 26 variable observations which are typically collected on an hourly basis. Once the raw data is submitted to a Samos server, an automated quality control software applies an initial set of flags on the data. This “preprocessing” step “(written in standard FORTRAN) was designed to flag data that failed to pass a series of objective evaluations.”
   3. The Problem
      1. Mimicking visual inspection of datasets.

* 1. A review of the relevant literature:
     1. (Neural networks used to classify)
        1. Robust Principle Component Analysis (Netflix)
        2. Seasonal Hybrid Extreme Studentized Deviate test (S-H-ESD) (Twitter)
        3. Regression using LSTMs, RNNs, DNNs or other

1. The project proposal:
   1. “performance of regression-based machine learning techniques on periodic time series data with limited features and few labeled examples of anomalies”
      1. periodic time series data - Seasonality is always of a fixed and known period. Hence, seasonal **time series** are sometimes called **periodic time series**.
      2. The samos observational data reflects a collection of observations related to seasonal and geospatial variables.
         1. Problem could be approached with a a regression-based machine learning.
      3. Detection
2. Potential Findings
   1. Training accuracy
   2. Test accuracy
   3. Comparative Analysis of regression methods
   4. Conclusion (Retrospective)