October 1, 2014
CS 74 – Machine Learning
Project Proposal

Location-Independent Weather Forecast

Andrew Pillsbury and Rebecca Leong

Overview:
Weather forecasting has traditionally been treated as a well-studied physics-based phenomenon for a specific location. Even so, weather still exhibits data patterns that can potentially be utilized as a basis for future prediction. By gathering data from a variety of different locations in the continental United States, we are attempting to create a location-neutral weather forecaster. For any given location, the model will predict 24 hours of future weather based on data from the past few days in that location. This will offer significant benefits over current methods of weather forecasting that require region-specific models.

Training Data:
The data we will use is quality controlled local climatological data from the National Climatic Data Center (http://cdo.ncdc.noaa.gov/qclcd/QCLCD?prior=N). We will use the last year of data from nine different cities: one for each climate region in the United States. These cities are: Seattle, WA, San Francisco, CA, Cortez, CO, Bismarck, ND, Dallas, TX, Atlanta, GA, Indianapolis, IN, Minneapolis, MN, Boston, MA. The weather stations record data roughly 40 times per day, so we anticipate having about 135,000 points of training data. Our training features will consist of weather data from every 4th hour in the 72 hours before T, where T is the beginning of the 24-hour period for which we wish to predict the weather. Each set of weather data will have measurements of visibility, temperature, dew point, relative humidity, wind speed, wind direction, station pressure, sea level pressure, and precipitation.

Testing Data:
The model will take the last 72 hours of weather data from a location of interest and predict the visibility, temperature, humidity, wind speed, and precipitation for that location in 4-hour intervals for the next 24 hours.

Algorithm:
The model will use a Dynamic Neural Network to make its predictions. Since the feature set will consist of 180 features (72/4 sets of 10 types of data), we will use feature selection to determine which features are the most relevant and reduce the size of the feature set. The Dynamic Neural Network will allow for pattern recognition of nonlinear multivariate behaviors while accounting for the time-dependence of the phenomenon, which will enable it to accurately and robustly predict upcoming weather.

Timeline:
By the milestone we hope to have a working prototype of our model so that we can begin training and testing to determine the optimal feature set, number of locations, and time-span of input data to use.

Sources:
Bontempi, Gianluca, Souhaib Ben Taieb, and Yann-Aël Le Borgne. "Machine learning strategies for time series forecasting." *Business Intelligence*. Springer Berlin Heidelberg, 2013. 62-77.
de Kock, Matthew. "Weather Forecasting Using Bayesian Networks." University of Cape Town. (2008)
Haerter, Fabricio P., and Haroldo Fraga de Campos Velho. "New approach to applying neural network in nonlinear dynamic model." *Applied Mathematical Modelling* 32.12 (2008): 2621-2633.
Lai, Loi Lei, et al. "Intelligent weather forecast." *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on*. Vol. 7. IEEE, 2004.

Rasouli, Kabir, William W. Hsieh, and Alex J. Cannon. "Daily streamflow forecasting by machine learning methods with weather and climate inputs." *Journal of Hydrology* 414 (2012): 284-293.

Shrivastava, Gyanesh, et al. "Application of Artificial Neural Networks in Weather Forecasting: A Comprehensive Literature Review." *International Journal of Computer Applications* 51.18 (2012): 0975-8887.

Sinha, N. K., M. M. Gupta, and D. H. Rao. "Dynamic neural networks: An overview." *Industrial Technology 2000. Proceedings of IEEE International Conference on*. Vol. 1. IEEE, 2000.

Soderland, Stephen. "Learning information extraction rules for semi-structured and free text." *Machine learning* 34.1-3 (1999): 233-272.

Williams, John K., et al. "A machine learning approach to finding weather regimes and skillful predictor combinations for short-term storm forecasting." *AMS 6th Conference on Artificial Intelligence Applications to Environmental Science and 13th Conference on Aviation, Range and Aerospace Meteorology*. 2008.