

Title: An Online Learning System for the Prediction of Electricity Distribution Feeder Failures

Authors: Hila Becker and Marta Arias, Center for Computational Learning Systems, Columbia University

Abstract:

We are using machine learning techniques for constructing a failure-susceptibility ranking of feeder cables that supply electricity to the boroughs of New York City. The electricity system is inherently dynamic, and thus our failure-susceptibility ranking system must be able to adapt to the latest conditions in real time, having to update its ranking accordingly. The feeders have a significant failure rate, and many resources are devoted to monitoring, maintenance and repair of feeders. The ability to predict failures allows the shifting from reactive to proactive maintenance, thus reducing costs.

The feature set for each feeder includes a mixture of static data (e.g. age and composition of each feeder section) and dynamic data (e.g. electrical load data for a feeder and its transformers). The values of the dynamic features are captured at the time of training and therefore lead to different models depending on the time and day at which each model is trained. Due to the seasonal change in electricity use which varies the load on the feeders, and other weather-related phenomena that may affect feeder failures, we have to account for a concept drift when designing our learning framework.

Previously, a framework was designed to train models using a new variant of boosting called Martingale Boosting, as well as Support Vector Machines. These models were trained using a snapshot of the data that was taken at a specific point in time (batch training) and then used to make predictions on future data as it became available. By constantly monitoring the performance of these models, we could observe that while the results were highly accurate at first, the performance would tend to decline as a function of the time between training and testing. Therefore, in this framework, an engineer had to decide whether to use the most recent data to build a new model, or use the latest model instead, for future predictions.

To avoid the need of human intervention, we have developed an “online” system that determines what model to use by monitoring past performance of previously trained models. In our new framework, we treat each batch-trained model as an expert, and use a measurement of its performance as the basis for reward or penalty of its quality score. Using this framework setting, we can leverage the promising performance that we previously observed in the batch-trained models, while dealing with the concept drift. The performance of the models is measured as a normalized average rank of failures. For example, in a ranking of 50 items with actual failures ranked #4 and #20, the performance is:

$$1 - (4 + 20) / (2 * 50) = 0.76.$$

Our approach builds on the notion of learning from expert advice as formulated in the continuous version of the Weighted Majority algorithm. Since each model is analogous to an expert and our system runs live thus gathering new data and generating new models, we have to keep adding new experts to the existing ensemble throughout the algorithm's execution. Every expert has an associated weight, which is used as a measure of confidence in its predictions. The prediction function of the algorithm combines the rankings of the top performing experts by taking a weighted average of each feeder's rank and sorting accordingly. To avoid having to monitor an ever-increasing set of experts, we drop poorly performing experts after each prediction.

We had to address the following key issues in our solution: (1) how often and with what weight do we add new experts, and (2) what experts do we drop. Our simulations suggest that using the median of all current models' weights for new models works best. To drop experts we use a combination of age of the model and past performance. Several other parameters of the algorithm which can be tuned to provide different tradeoffs include the number of models used to make the final prediction, the maximum allowable ensemble size and the learning rate.

Our system is currently deployed and being tested by New York City's electricity distribution company. Results are highly encouraging, with 75% of the failures in the summer of 2005 being ranked in the top 26%, and 75% of failures in 2006 being ranked in the top 36%.