

Estimating the Time Between Failures of Electrical Feeders in the New York Power Grid

Haimonti Dutta, David Waltz, Alessandro Moschitti, Daniele Pighin, Philip Gross, Claire Monteleoni, Ansaf Salieb-Aouissi, Albert Boulanger, Manoj Pooleery and Roger Anderson
Center for Computational Learning Systems, Columbia University, New York, NY 10115
{haimonti, waltz}@ccls.columbia.edu; moschitti@disi.unitn.it;
daniele.pighin@gmail.com; {phil, cmontel, ansaf, aboulanger, manoj, anderson}@ccls.columbia.edu

1 Introduction

Electricity generated by steam or hydro-turbines at power plants is transmitted at very high voltage from generating stations to substations, distributed from substations to local transformers via *feeder* cables and finally sent to individual customers [8] (Figure 1). In the New York City Power Grid, a little more than 1000 *primary distribution feeders* transmit electricity between the high voltage transmission system and the household-voltage secondary system. These feeders are susceptible to different kinds of failures such as emergency isolation caused by automatic substation relays (Open Autos), failing on test, maintenance crew noticing problems and scheduled work on different sections of the feeder.

Over the past few years, researchers at CCLS have collaborated with the Consolidated Edison Company of New York to develop systems that can *rank* feeders and their components (cable sections, joints and splices) according to their susceptibility to failure. The Ranker for Open-Auto Maintenance Scheduling (ROAMS) [8] was the first such system built using Martingale Ranking [14]. Subsequently, the system was improved to boost ranking performance using an ensemble of ranking experts [2] which, however, came at a cost of interpretability of the machine learning models. A comparison of three different techniques for ranking electrical feeders - Martingale Ranking, RankBoost and SVM score ranker can be found in [9]. More recently, we have begun to focus on estimating measures such as Time Between Failures (TBF) of feeders resulting in regression problems as opposed to ranking. In this paper, we describe the challenges faced, our approach to modeling the problem and provide empirical results obtained from the models.

This paper is organized as follows: Section 2 presents related work; Section 3 the challenges faced, Section 4 the data generation process, Section 5 our approaches to modeling Time Between Failures (TBFs) and Section 6 presents a case study for modeling TBFs of feeder cables in Brooklyn and Queens.

2 Related Work

Modeling failure rates (i.e. frequency with which an engineered system or component fails) has been studied extensively in reliability theory ([7], [11]). Begovic et al. [3] study parametric statistical models when only partial information (such as installation date, number of components replaced in a given year, failure and replacement rates) is available. They use Weibull distributions to model future failures and for formulating replacement strategies. The Cox Proportional Hazards model [6] is another semi-parametric regression where the features are modeled as scaling the instantaneous failure rate

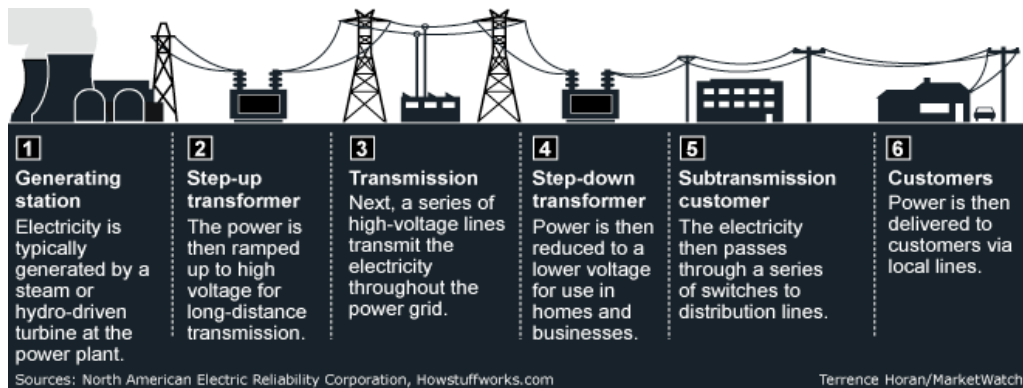


Figure 1: Electricity Generation and Distribution

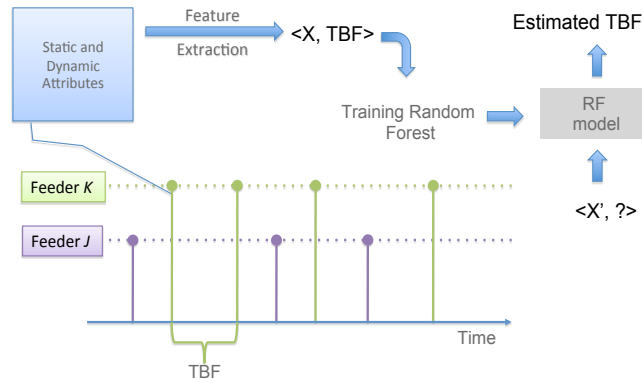


Figure 2: **The Sampling Procedure for collecting data.**

of a component. Guo et al [10] propose a model based on Proportional Intensity and explore tools to analyze repairable systems – their approach can incorporate time trends, proportional failure intensity and cumulative repair effects. A technique for approximating the mean time between failure of a system with periodic maintenance is described in work done by Mondro [15] while models for recurrent events are studied by Lawless and Thiagarajah [12]. In this paper, we present challenges faced and preliminary results in estimating TBFs using *machine learning* techniques in the New York power grid.

3 Challenges in Estimating Time Between Failures

Several challenges exist in generating good regression models for estimating Time Between Failures (TBFs):

1. Few components actually failed during the time for which we have data. Many components have never failed or failed only once during the time for which we have data, and for these cases we need to learn estimates from “censored” data, i.e., data on time intervals where we only know that TBF is greater than a) the period for which we have data, or b) the times from the last failure before we started collecting data until the first failure, or c) the time from the last failure until the present. In some cases we may have two or more failures of the same feeder during the collection period, making it possible to get more precise data to train on.
2. Another key challenge is that there are several failure modes (such as Open Autos, Failed on Test, Out on Emergency), and so task of prediction is highly non-linear. Key failure causes for feeders include aging, power quality events (e.g. spikes), overloads (that have seasonal variation, with summer heat waves especially problematic), known weak components (e.g. PILC cable and joints connecting PILC to other sections), at-risk topologies (where cascading failures could occur), workmanship problems and the stress of HiPot testing and deenergizing/reenergizing of feeders.
3. Since there are many different causes of failures, it is difficult to pin-point an exact cause of failure; furthermore the same feeder can fail multiple times within a short time span (often called “infant mortality”) or last more than a few years. Thus there are considerable fluctuations in survival times resulting in a very imbalanced data set.
4. As pointed out by Begovic et al. [3], “An accurate model of power apparatus lifetime should contain a large number of factors, which are not practical for monitoring - a partial list should contain the initial quality and uniformity of the materials the equipment is made of (primarily the insulation), the history of exposure to moisture, impulse stress, mechanical stress, and many other factors. As those are neither available in typical situations (databases often do not even associate failures with the age), nor is their impact well documented and understood, the model that captures the essential behavior is, by necessity and for practical reasons, chosen to contain the most salient features known to be the strong determinants of lifetime.”
5. Sensors on equipment capture long time series data such as current load on a feeder, power quality events and other composite measurements of stress on the feeder. This results in creation of huge asynchronous time series databases – aggregation, interpolation and mining of which provides formidable challenges.

4 Data Generation

Snapshots of the state of a feeder are taken at the time of the failure (Figure 2). For each feeder, the attributes comprise of: (a) physical characteristics (such as number of cable sections, joints, installed shunts); these characteristics may undergo

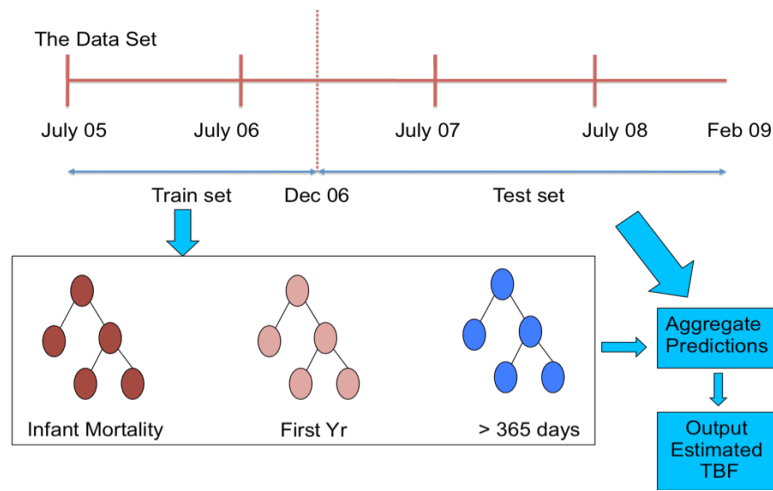


Figure 3: **Random Forests for modeling Time Between Failures (TBFs) of Electrical Feeders.**

annual changes. (b) electrical characteristics from load flow simulations (c) dynamic data from telemetry attached to the feeder (such as power quality events, load forecasts, outage counts) and (d) derived attributed suggested by domain experts. Since the number of feeders vary considerably in different boroughs in New York, the size of the data set on which machine learning models are built is also different. There are approximately 900 instances in the data set for Manhattan, 1300 for Brooklyn and Queens and 350 for Bronx. The training and validation data was collected from July, 2005 - December 2006 and blind-testing is done on data from January 2007 – February 2009.

5 Modeling Time Between Failures (TBFs)

We have applied Support Vector Machines (SVM) [16], CART (Classification and Regression Trees) [5], ensemble based techniques (Random Forests [4]) along with statistical methods, e.g. Cox Proportional Hazards [6], to the task of estimating TBF. In this paper, we present empirical results obtained from Random Forest (RF) based models (illustrated in Figure 3). To effectively model short time survivors¹ (feeders which have failures within 90 days of the last failure) and long time survivors (feeders surviving greater than 365 days), we built classification and regression trees for each class of survivors (short survivors, one year survivors and long term survivors). Our hypothesis was that similar kinds of failures should be modeled using a single regression model. The decision of how to partition the TBFs for building different models was based on knowledge acquired from domain experts. However, for the infant mortal cases, i.e. whether to make a model for feeders that failed within 10 days versus 20 days, we relied entirely on empirical analysis.

Once the Random Forest of trees was generated, the next problem was how to *combine* the regression models such that when an unseen test example was presented to it, we could accurately predict the time between failures. The process of combining models was tricky because different models predicted TBFs in different ranges and simply averaging results was unreasonable; for instance say a test instance was passed through an RF model and regressors for each class were allowed to come up with predictions; assume the short survivor model predicted a TBF of 2 days, the one year survivor predicted 265 days and the long term survivor predicted 1089 days; an average yields 452 days which does not indicate whether the feeder is generally one that is infant mortal or a longer survivor or neither. To avoid this problem, we tried different mechanisms of combining models: (1) Weighted Averaging, where weights were the proportion of instances in the training set that belong to a particular class of survivors (2) Build a decision tree on the training data to obtain class labels – 0 indicating short survivor, 1 representing one year survivors and 2 representing long term survivors. Given a test instance, the decision tree predicts which class it belongs to and then we use the corresponding tree to come up with a prediction. (4) Clustering the training data using (a) K-Nearest Neighbor (b) K-Means and (c) KMeans++ each with three different distance metrics – euclidean distance, L1 Norm and the cosine metric. We also investigated different seeding mechanisms for initiating the clustering algorithms. “Seeding” a clustering algorithm is the problem of choosing the initial cluster centers as input to an algorithm. For example, the canonical K-means algorithm, takes as input a set of k “seeds” which are the first set of candidate centers to the iterative algorithm. A recent, significant advance in k-means clustering seeding technology was made by Arthur and Vassilvitskii ([1]). Their algorithm, k-means++, is extremely light-weight and simple, however has strong formal performance guarantees (the clustering, induced by the seeding alone, approximates the optimum value of the k-means clustering objective by a factor of $O(\log k)$). We implemented this procedure to seed the clustering of TBFs, along the time dimension, for the purposes of finding good 3-clusterings. In the following section

¹These are also referred to as instances suffering from “infant mortality” (IM)

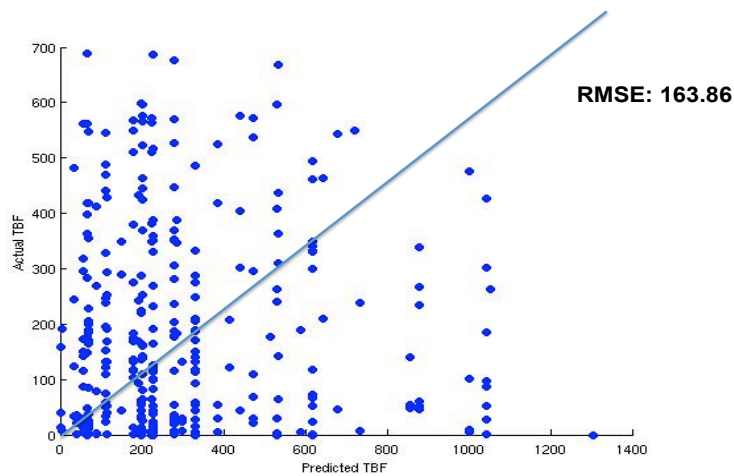


Figure 4: Actual vs Predicted Time Between Failures on the Brooklyn and Queens Data using a Random Forest of three trees and combination by k-Nearest Neighbors method.

we present empirical results for estimating TBFs in Brooklyn and Queens².

6 Case Study: Estimating TBFs of Feeders in Brooklyn and Queens

The training data collected between July 2005 - July 2006 contains approximately 1100 instances. This data is used to construct three different regression trees corresponding to short survivors, one year survivors and long term survivors. Since this a real-world application and defining what comprises a “short” survivor is non-trivial, we built models for feeders that failed in 10 days, 20 days, . . . , 90 days³. When we built a 10 day short survivor model, the one year model was built on all instances greater than 10 days and less than 365 days and the long term model contained instances that survived longer than 365 days. These models were first tested on a validation data collected between August 2006 - December 2006. In addition to building the regression models on the training data, we also constructed machine learning models (decision trees, clusters) for *combining* regression models as described in Section 5. The blind testing was done on data collected between January 2007 – February 2009. The metric used to evaluate our models is the Root Mean Square Error (RMSE)⁴. Table 1 shows the RMSE values of the Random Forest models with different model combination techniques in Brooklyn and Queens. Our results indicate that the best model (lowest RMSE value) is obtained when trees are built by splitting instances that fail within 80 days of the last failure as the short survivors, instances corresponding to failures greater than 80 days and less than 365 days as one-year survivors and those greater than 365 days as long term survivors and combining them using weighted averaging. However, the predictions of such a model tend to be restricted to within a year of the last failure. To avoid this and allow a larger range of predictability, we use the k-Nearest Neighbor model with euclidean distance measure. Figure 4 shows the plot of actual TBF versus predicted TBF for this model. Our results indicate that we are able to predict Time Between Failures within approximately 6 months from the last failure. Future work includes incorporation of seasonal trends in the model and use of other machine learning techniques to build more robust random forests.

Acknowledgements

This work is supported by the Consolidated Edison Company of New York.

References

- [1] David Arthur and Sergei Vassilvitskii. k-means++: the advantages of careful seeding. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

²The results from other boroughs are left out for the sake of brevity.

³This was suggested by domain experts.

⁴The Mean Squared Error (MSE) measures the average of the square of the “error” where error is the amount by which the true value differs from the quantity estimated. The square root of the MSE yields RMSE which is the metric used to evaluate our models.

Combination Method	10d	20d	30d	40d	50d	60d	70d	80d	90d
Averaging	196.35	191.46	193.32	196.45	210.09	201.8	217.07	209.74	218.13
Wt Avg	146.49	133.07	131.85	129.37	139.23	127.19	137.07	126.55	131.78
Decision Tree	251.87	312.26	263.86	268.96	314.87	266.99	276.1	270.86	263.27
k-NN Euclidean	190.7	163.86	168.15	172.42	201.35	181.55	193.43	179.29	183.39
k-NN L1 Norm	187.14	170.26	176.33	179.57	195.83	175.45	192	183.13	184.05
k-NN Cosine	195.37	191.71	190.85	197.88	215.6	204.34	217.29	201.46	202.82
k-means Euclidean	244.51	243.75	243.96	239.29	249.93	247.28	251.87	245.39	257.83
k-means L1 Norm	274.47	271.81	269.81	269.45	276.72	322.58	327.6	322.53	282.59
k-means Cosine	269.06	265.2	267.62	259.67	280.36	274.83	281.28	272.49	255.98
k-means++ Euclidean	244.51	242.32	243.31	239.29	249.93	243.05	251.13	245.23	246.76
k-means++ L1 Norm	195	185.56	183.95	183.03	201.84	194.05	202.73	189	203.88
k-means++ Cosine	230.26	223.75	224.19	226.06	228	225.37	227.59	222.25	226.84

Table 1: **RMSE values for Random Forest models built on Brooklyn and Queens data.**

- [2] Hila Becker and Marta Arias. Real-time ranking with concept drift using expert advice. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD '07)*, pages 86–94, New York, NY, USA, 2007. ACM.
- [3] M. Begovic, P. Djuric, J. Perkel, B. Vidakovic, and D. Novosel. New probabilistic method for estimation of equipment failures and development of replacement strategies. *Hawaii International Conference on System Sciences*, 10:246a, 2006.
- [4] Leo Breiman. Random forests. In *Machine Learning*, volume 45, pages 5–32, 2001.
- [5] Leo Breiman, Jerome Friedman, Charles Stone, and R Olshen. *Classification and Regression Trees*. CRC Press, Boca Raton, Florida, USA., 1998.
- [6] D. R. Cox. Regression models and life tables. In *J. Roy Statist Soc B*, volume 34, pages 187 – 220, 1972.
- [7] Charles E Ebeling. *An Introduction to Reliability and Maintainability Engineering*. McGraw-Hill Companies, Inc., Boston, USA., 1997.
- [8] P. Gross, A. Boulanger, M. Arias, D. L. Waltz, P. M. Long, C. Lawson, R. Anderson, M. Koenig, M. Mastrocinque, W. Fairechio, J. A. Johnson, S. Lee, F. Doherty, and A. Kressner. Predicting electricity distribution feeder failures using machine learning susceptibility analysis. In *The Eighteenth Conference on Innovative Applications of Artificial Intelligence IAAI-06*, Boston, Massachusetts, 2006.
- [9] Phil Gross, Ansaf Salleb-Aouissi, Haimonti Dutta, and Albert Boulanger. Ranking electrical feeders of the new york power grid. In *3rd Annual Machine Learning Symposium at the New York Academy of Sciences (NYAS)*, New York, NY, October 2008.
- [10] W Zhao H Guo and A Mettas. Practical methods for modeling repairable systems with time trends and repair effects. *Proceedings of Annual Reliability and Maintainability Symposium*, 2006.
- [11] K. C. Kapur and L. R. Lamberson. *Reliability in Engineering Design*. John Wiley and Sons, New York, USA., 1977.
- [12] T. F. Lawless and K. Thiagarajah. A point-process model incorporating renewals and time trends with applications to repairable systems. *Technometrics*, 38(2), 1996.
- [13] S. P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
- [14] Philip M. Long and Rocco A. Servedio. Martingale boosting. In *Learning Theory: 18th Annual Conference on Learning Theory, COLT 2005, Bertinoro, Italy, June 27-30, 2005, Proceedings*, volume 3559 of *Lecture Notes in Artificial Intelligence*, pages 79–94. Springer, 2005.
- [15] M. J. Mondro. Approximation of mean time between failure when a system has periodic maintenance. *IEEE Transactions on Reliability*, 51(2), 2002.
- [16] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.