Predicting Marginal Generators in Real Time

Priya L. Donti,[*,1,2] Inês Lima Azevedo,[1] and J. Zico Kolter[2]

[1] Engineering & Public Policy, Carnegie Mellon University; [2] Computer Science, Carnegie Mellon University

[*] pdonti@cmu.edu

Introduction: Marginal Generators

Real-time characterization of marginal generators is important for power system operation and emissions reduction.

Marginal Generators: The last generator(s) needed to meet electricity demand at a given time, and the first to be affected by an intervention.

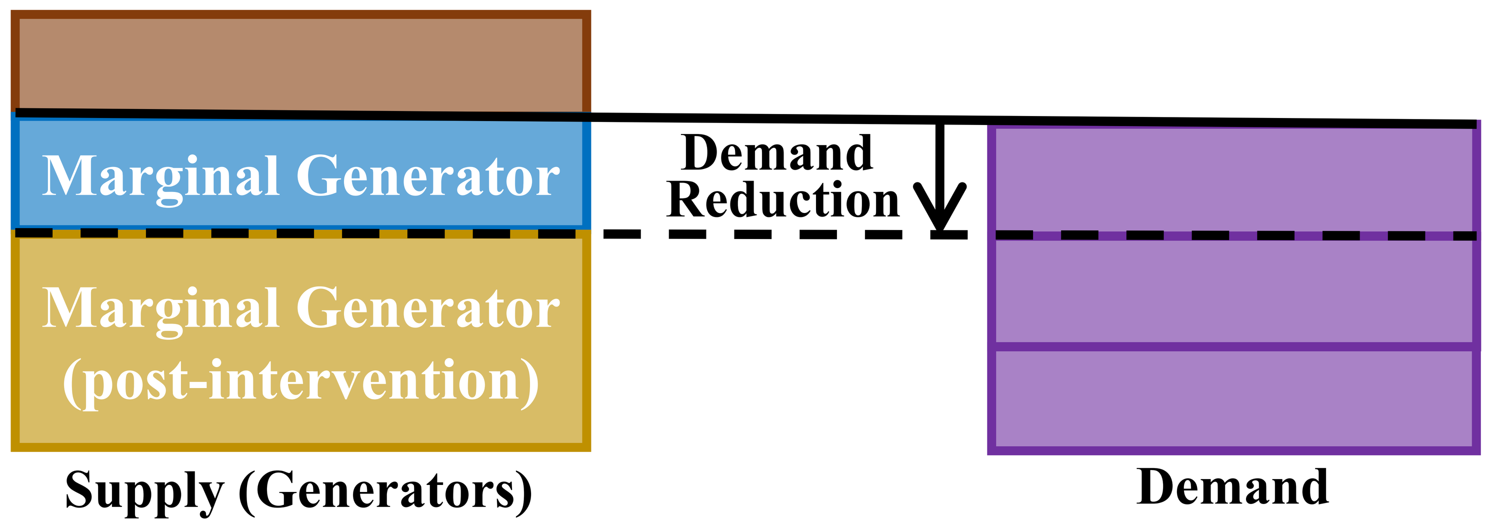

Figure 1: Simplified illustration of marginal generation at a given time. An electricity demand-reducing intervention (e.g.) might displace the original marginal generator.

Marginal generation is typically predicted using dispatch models, which can be technically and managerially complex.

We instead employ machine learning models on historical data (as in [1] and [2]) **to predict marginal generators in near-real-time.**

Historical Data

We constructed our prediction models using historical data publicly available from PJM and other sources.

- Hourly marginal fuel types and proportions (aggregated from the five-minute level) [PJM] *(See Figure 2)*
- Hourly metered loads and load forecasts [PJM]
- Real-time and day-ahead hourly integrated LMPs [PJM]
- Real-time and day-ahead hourly transmission constraints [PJM]
- Temp., wind speed, pressure, humidity, dew point [Wunderground]
- Coal and natural gas electric power prices [EIA]

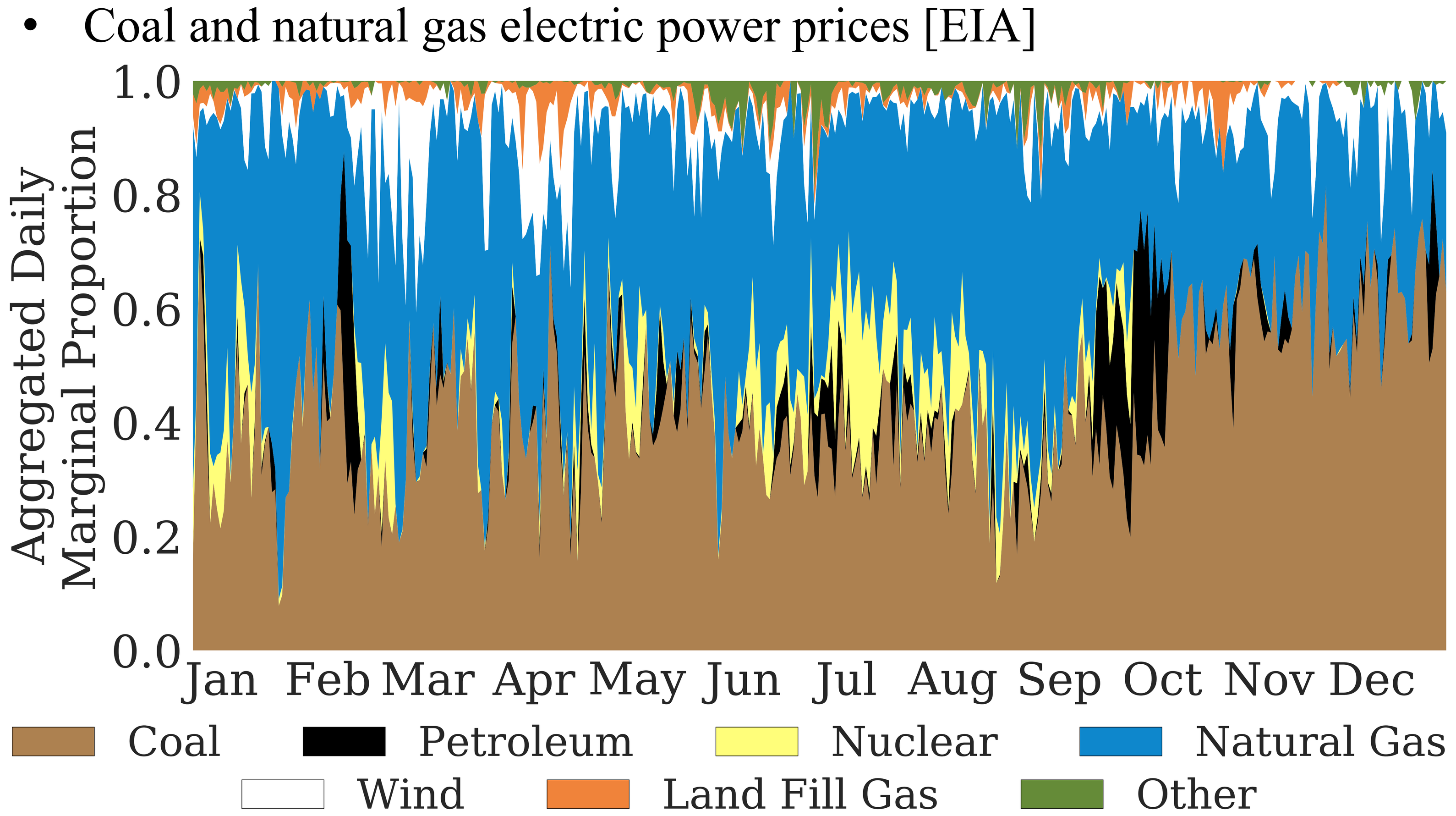

Figure 2: Time series of 2012 PJM marginal fuel proportions, aggregated by day. Coal and natural gas are the most common marginal fuels.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE1252522.

Method: Real-Time Classification

We employed three modified gradient boosting classification algorithms to predict each fuel type's hourly marginal proportion.

Algorithm Overview: Three classification algorithms using different combinations of objective functions and evaluation errors.

- Inputs: Unique feature sets for each hour, constructed from the perspective of a planner making forecasts two hours before the given time point (to reflect lags in data availability) *(See Table 1)*
- Outputs: Hourly marginal proportions for each of 7 fuel types (Coal, Natural Gas, Petroleum, Land Fill Gas, Nuclear, Wind, Other)
- Objective functions/evaluation errors: Softmax (normalized exponential), total variation (L_1), and/or least squares (L_2)

Table 1: Features included for every time point in our model. RT and DA indicate "real time" and "day ahead," respectively.

Feature	Information Included	Area(s)/Zone(s)
Fuel Mix	- 2-4 hr prev. - Same hr prev. day	PJM-wide
Metered Load	- 2-4 hr prev. - Same hr prev. day - Total load prev. day	All control, NERC, market, and load zones excl. EKPC (24)
Load Forecast	- Given hr - Total for current day	All forecast zones (9)
LMPs & RTO Price	- RT: 2-4 hr prev. - DA:Given hr	All aggregate zones (22) and PJM-wide
Transmission Constraints	- DA: Given hr	PJM-wide
Weather	- Given hr - 1-2 hr prev.	Major PJM-region airports (9)
Natural Gas Price	- Current monthly price (or last avail.)	U.S. & Pennsylvania
Coal Price	- Current quarterly price (or last avail.)	U.S. & Mid-Atlantic
Time	- Hr - Weekday - Mo. - Weekend?	[N/A]

Algorithm Details:

- Implemented using the XGBoost library [3]
- Train on 2011-14 data, hold out 2015, test on 2016
- Input one hourly sample per marginal fuel type category, with fuel type as class label and marginal proportion as sample weight
- Run training algorithm until evaluation error on holdout set has not decreased in ten training "rounds"
- Evaluate predictions on testing set. Algorithm outputs ordinal probabilities that each testing sample is in each fuel class, which we interpret as cardinal probabilities when evaluating testing error.

Preliminary Results

We were able to predict marginal generation with reasonable accuracy and characterize factors important to our predictions.

The errors obtained by our models on the test data set (2016) are reported in Table 2, alongside baselines for comparison.

- L_2 evaluation error models both outperform 4/5 baselines
- Softmax objective/L_1 evaluation error model beats 3/5 baselines
- Previous-hour, median, and mean baselines reflect data not yet available at prediction time

Table 2: Model prediction errors for 2016 data, compared to always guessing the median or mean marginal fuel proportions, or the marginal proportions from one hour prior to each time point, two hours prior, or the same hour the previous day.

	Model Objective		Baseline				
Eval. Error	*Softmax*	L_2	*Median*	*Mean*	*Prev. Day*	*2 Hr Prev.*	*Prev. Hr*
L_1	0.194	-	0.208	0.247	0.233	0.180	0.147
L_2	0.072	0.081	0.109	0.101	0.127	0.086	0.059

Key predictive features (by XGBoost's "feature importance") were:

- All models: Zonal LMPs, prior marginal fuel proportions, transmission constraints
- Softmax objective models: Natural gas prices, wind speed
- L_2 objective models: Zonal metered loads

Table 3: Top 10 most important features in both models employing L_2 evaluation error. Zonal LMPs and recent marginal fuel proportions were important in both cases.

Softmax Objective	L_2 Objective
U.S. NG Electric Power Price	AE-DayMeteredLoad-yest
Coal-2HR	AE-HrMeteredLoad-2HR
Natural Gas-2HR	DEOK-TotalLMP-DayAhead
RTO-CongLMP-DayAhead	Coal-2HR
TransConstraints-DayAhead	Light Oil-2HR
RTO-LossLMP-DayAhead	AE-HrMeteredLoad-yest
DEOK-TotalLMP-DayAhead	Natural Gas-2HR
DPL-CongLMP-DayAhead	Wind-2HR
HrWindSpeed-MDW	Nuclear-2HR
DOM-LossLMP-DayAhead	TransConstraints-DayAhead

Summary and Conclusions

- We present a machine learning-based method to predict the fuel types of marginal generators in near-real-time using historical data.
- **Initial results seem promising, with test errors comparing favorably to even somewhat prescient baselines.**
- Future work includes characterizing the emissions profiles suggested by our predictions; developing methods that inherently treat proportions as cardinal (not ordinal) concepts; and evaluating the implications of different objective functions (ongoing work).

References

[1] K. Siler-Evans, I. L. Azevedo, and M. G. Morgan, "Marginal Emissions Factors for the U.S. Electricity System," Environmental Science and Technology, vol. 46, no. 9, pp. 4742–4748, 2012.

[2] C. Wang, Y. Wang, C. J. Miller, and J. Lin, "Estimating Hourly Marginal Emission in Real Time for PJM Market Area Using a Machine Learning Approach," in Proc. of the Power and Energy Society General Meeting (PESGM), pp. 1–5, 2016.

[3] T. Chen and C. Guestrin, "Xgboost: A Scalable Tree Boosting System," in Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.