Forecasting Coincident Peaks with a Feed-Forward Neural Network

Chase Dowling, Daniel Kirschen, Baosen Zhang

University of Washington,
Electrical and Computer Engineering

November 28, 2018

Coincident Peak Charging

An electrical customer's coincident peak (CP) is their demand at the moment of the entire system's peak

Systems levy transmission surcharges via CP electrical rates to reduce system peaks.

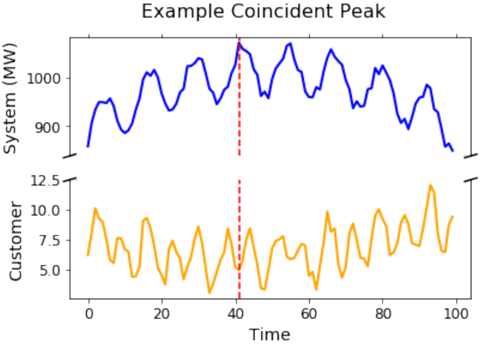


Figure 1: Example CP

How do CP charges work?

1. CP rate roughly 100x more than normal time-of-use rates
2. A consumer's CP is recorded on a monthly basis
3. At the end of the year, CP charges are paid

Consumers participate in exchange for discounted time-of-use rates at all other times.

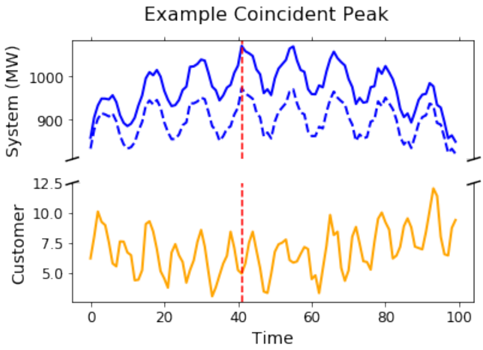


Figure 2: Example of idealized system response to CP charges

Motivation

4 MW consumer paying average ERCOT wholesale prices ($40/MWh), roughly $1.4 million in electricity costs per working year, $300k of which per year to consume electricity at CP hour
Consumers are incentivized to *curtail demand during the moment of the CP*.

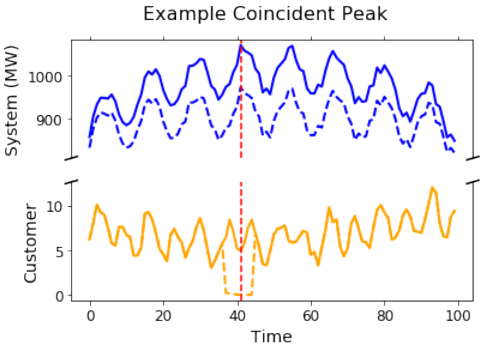


Figure 3: Example of a consumer's CP

Current Solutions

Operators broadcast signals, e.g. Fort Collins PUD:

- Sends out signals about 10 days out of month
- Signals come with less than one hour lead time
- Customers know when CP's *should* occur, e.g. afternoon

Too many signals, still hard to predict rare events

Distributors and Large Consumers

Contribution

1. Cast the CP prediction problem in the context of an optimization program a consumer can evaluate
2. Treat the CP occurence as a random variable
3. Design a context-aware loss function for training predictors in this regime

Optimization Problem

Let g be a concave/differentiable utility function of hourly power consumption p_t, π_{cp} the CP rate: we have the following optimization program:

$$\begin{aligned} \underset{p_t}{\text{maximize}} \quad & g(p_t) - \pi_{cp} p_t \cdot \mathbf{1}[t \text{ is CP}] \\ \text{subject to} \quad & p_t \leq p_{\max} \\ & p_t \geq 0 \end{aligned} \tag{P1}$$

CP as Random Variable

Instead of predicting binary sequences for moment that is/isn't a rare CP, we can instead treat the system load as a continuous RV try to predict the CDF

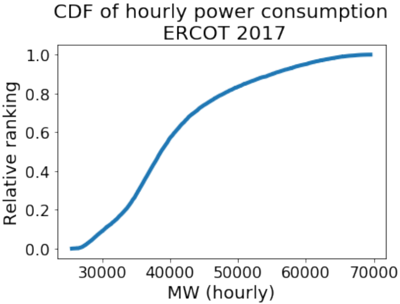


Figure 4: Emprical CDF of 2017 ERCOT hourly system load

Optimization Problem Relaxation

Now we can replace the indicator function with a probabilistic expression we can more effectively predict and take advantage of via thresholding:

$$\begin{aligned}\underset{p_t}{\text{maximize}} \quad & g(p_t) - \pi_{cp} p_t \cdot \mathbf{1}\left[CDF(p_t) \geq \alpha\right] && \text{(1a)}\\ \text{subject to} \quad & p_t \leq p_{\max} && \text{(1b)}\\ & p_t \geq 0 && \text{(1c)}\end{aligned}$$

Optimization Problem Relaxation

If the predicting CDF value is greater than some threshold α, then the CP cost factors into the optimization program.

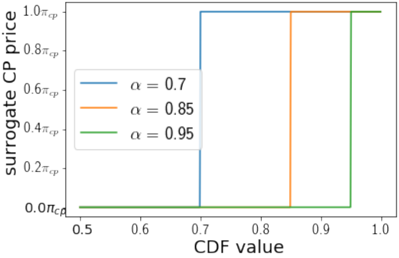


Figure 5: CP cost as a function of the predicted CDF value for various values of α

Optimization Problem Relaxation

This is still too stringent, so we relax curtailment severity and "hedge" our bet the upcoming demand-hour is a CP, curtailing more the larger the predicted CDF value is.

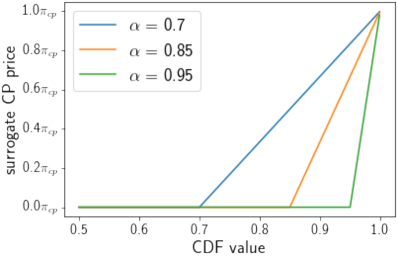


Figure 6: Hedged CP cost as a function of the predicted CDF value for various values of α

Predicting the CDF

We know weather is a major factor.

Predicting the CDF

We know weather is a major factor.

But we care about a large number of features that will influence small differences at the extremes.

Predicting the CDF

We know weather is a major factor.

But we care about a large number of features that will influence small differences at the extremes.

NN's are good at incorporating a large number of features. We highlight that a simple, linear NN can identify and exploit such features for the purpose of more accurately predicting the timing of the system peak. The model is learning something.

Predicting the CDF

We know weather is a major factor.

But we care about a large number of features that will influence small differences at the extremes.

NN's are good at incorporating a large number of features. We highlight that a simple, linear NN can identify and exploit such features for the purpose of more accurately predicting the timing of the system peak. The model is learning something.

2017 is an interesting test year: ERCOT peak loads were roughly 4,000 MW less than predicted values that hour-ahead forecasts would be benchmarked against.

Predicting the CDF

We want to promote prediction accuracy for larger values of the CDF, between $[\alpha, 1]$, so we design a weighted average L1 loss

$$\mathcal{L}_\beta := \frac{1}{|\{P\}|} \sum_{x_t \in \{X\}} \left[\beta^{F(S_t)} |F(S_t) - \hat{F}(S_t)| \right] \tag{2}$$

With this loss function we can frontload a NN with feature data to predict the CDF, and test the effectiveness of the predicted CDF in our consumer's optimization program.

ERCOT Load Data

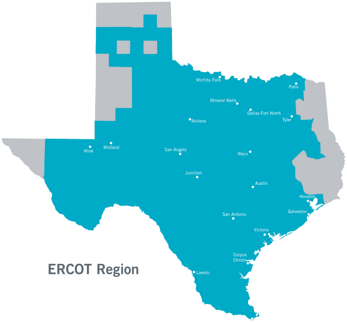


Figure 7: Map of ERCOT region and major metropolitan areas

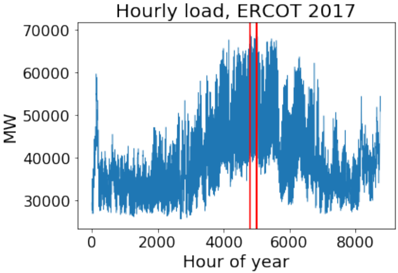
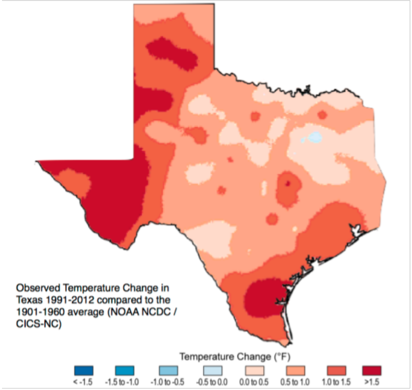


Figure 8: Historical ERCOT system load data 2017

Weather Data

- Hourly data from 2010-2017 retrieved from Dark Sky API
- Polled 19 largest cities
- Feature rich: temperature, humidity, precipitation, wind velocity, barometric pressure, cloud cover, etc.



Source: NOAA

Feature Engineering

We incorporate on a per-zone basis load features to account for potential congestion affects, and the non-uniform influence of weather across Texas. For example:

- Day of month, week, year, time of day, season
- ERCOT demand *by zone*
- Average, max, and variance of electrical demands observed thus far in the CP measurement period
- Average, max, and variance of temperature, humidity, and wind speeds observed thus far

Loss Function

Identical networks trained with standard average L1 loss and weighted L1 loss

Average L1 loss on test data—post training—for large values of α strictly improved

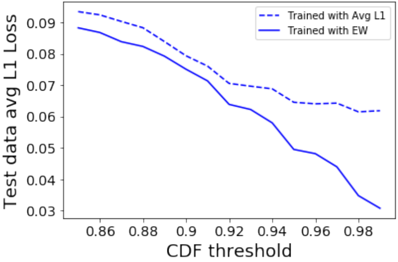


Figure 9: Comparison of avg. L1 and weighted avg. L1, $\beta = 10$

CP Identification Results

Make the problem harder by identifying top 10 loads annually. Use historical average model based on how Fort Collins PUD broadcasts expected CP time ranges.

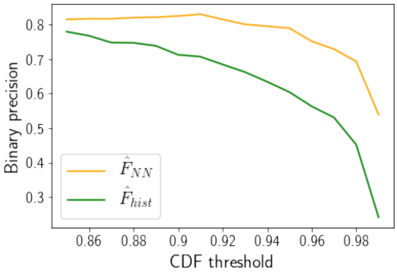


Figure 10: Binary Precision

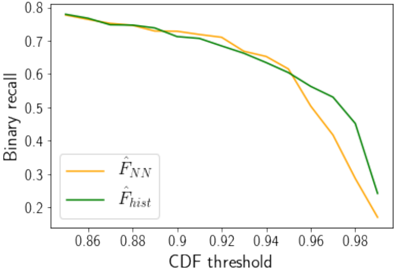


Figure 11: Binary Recall

Model Business Optimization Results

500 MW toy business, 10 CP charges over the entire year. Unit utility per MW during regular business hours, CP charge approx. 40% of annual utility.

1. 24 hour ahead NN prediction (97.4% of perfect)
2. Historical average CDF (94.4% of perfect)

Utility maximized for for 2 and 3 at $\alpha = 0.975$

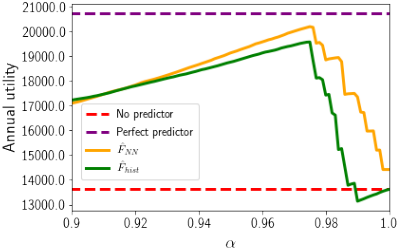


Figure 12: Comparison of CP curtailment strategies

Concluding Remarks

For small consumers, hedged CP curtailment can save considerable CP costs.

Simple predictor learns something over historical empirical CDF

Large consumers might change timing of system CP if they curtail enough. Interesting game theory problem.