Estimating marginal local energy prices using Deep Learning techniques

Blosher Brar (11258419), Melany Delgado (11249139), Ernest Tafolong (11119080)
MATH 80600A - Machine Learning II: Deep Learning J02
HEC Montreal, Department of Decision Sciences

1 - Introduction

The buying and selling of energy are common transactions in the Canadian and American markets. Indeed, since the electricity produced is difficult to store, it is important to produce enough to meet market demands, but not in excess to avoid large price drops that could lead to losses for electricity producers and transporters. The price of electricity varies depending on the geographic point where it is calculated. Indeed, it depends on the electricity demand, the production capacity of the suppliers (hydro-electric, solar, wind, gas, oil, etc.), normal losses of electricity caused by the network, and constraints such as events related to equipment shutdown or malfunction. This local price or LMP (local marginal price) is calculated twice: First at the day-ahead-market (DA) and then at the real-time market (RT). In the DA market, the independent system operator (ISO) collects buy and sell offers from the market participants and calculates the LMP for each hour the next day based on expected network conditions (Matherm and al. 2017). The RT price is calculated according to the actual system conditions. This structure of two markets, as well as the competition created by virtual transactions, results in price convergence and improved market efficiency, which translates into increased network reliability and lower costs for consumers (Baltaoglu et al. 2015).

Several techniques are used to predict electricity prices: Fourier transformations, autoregression models, Machine Learning (ML), and Deep Learning (DL) among others (Weron 2014). However, it is not possible to have a generalized method for all markets because each market has unique geographic, meteorological, and demographic characteristics.

During this project, we wanted to explore the application of deep learning models, including multilayer perceptron, convolutional networks, and recurrent networks to predict DA and RT prices to estimate DART for the PJM market. Our goal was to develop, and test models that estimate prices for the two markets. Also, since there are 49 transaction points (points that postpone an LMP), it was also a question of choosing the transaction point that would be the most profitable.

This report is organized as follows: first, the problem will be explained in more depth and a review of the relevant literature will be made, followed by a description of the different sources of the data and the models used. Next, the implementation of the selected models will be described, followed by the presentation of the results of the models and their analysis. Finally, other opportunities to explore will be discussed.

2 - Problem Description

Pennsylvania – New Jersey – Maryland (PJM) interconnection is an ISO that coordinates the distribution and delivery of electricity in 13 states of the United States as well as the District of Columbia. PJM serves 65 million people over an area of 84,236 square miles. (PJM interconnection, 2019). PJM has adopted the structure of two markets (DA and RT) for its energy market.

As already explained, an inherent problem with the electricity production business is that electricity cannot be easily stored once it is produced, and the demand for electricity fluctuates daily depending on very broad conditions, such as weather, temperature, economic activity, and general randomness associated with individual energy consumption. As such, electricity generating utility companies need to produce enough energy to satisfy the demand for electricity, while getting the best export price for the excess electricity generated. The export price in-turn depends on the balance between supply and demand in the electricity market. If there is more supply than demand, the price will fall, and if there is more demand than supply, the price will rise. Subsequently, if a utility firm is able to predict the future price of electricity at time t + 1, it can compare it to the current price at time t, and determine whether it should sell or hold its electricity. Moreover, it can even “virtually” buy electricity at time t and then sell it back to the market at time t + 1 to make a profit. Having the ability to accurately predict electricity prices has the potential to generate a significant profit for firms involved in the electricity market. The methods that are currently used for this purpose involve the use of historical data, demand predictions, and general knowledge of the electricity market to estimate the DART (Day After price and Real Time price). This project will aim to improve the accuracy beyond the currently used methods using Deep Learning techniques.

3 - Related Work

On the electricity market, performing day-ahead and real time prices that are consistent with the relationship between electricity supply/demand is critical to make informed decisions. For example, From the demand side, some companies can schedule their operations according to the low-price zones and operate in these hours or months. However, electricity usually exhibits complex features (non-linearity, non stationary, price spikes, seasonality, etc…), which render price’s prediction very challenging. Using statistical and machine learning methods is commonly used in the forecasting literature (Weron et al. 2014)

To address this issue and predict the nonlinear behavior of hourly prices, different machine learning methods have been proposed. Among them, multilayer perceptron (MLPs) (Szkuta et al. 1999). LSTM network has been widely used in various applications such as natural language processing and time series analysis. It is capable of learning features and long term dependencies of the historical information on the current predictions for sequential data (Sutskever, Vinyals, and Le 2014). RNN structures and LSTM have shown to be a much better alternative to accurately forecast complex time series. Very recently, L. Jiang and G. Hu used LSTM to accurately forecast the electricity in Victoria (Australia) and Singapore. Similarly, the

LSTM deep neural network has been proven to outperform alternative models on electricity predictions (Zhu et al. 2018).

Deep feature selection algorithms (Feng et al. 2017) and ensemble methods based on CNN has been proposed to accurately forecast in the energy area (Wang et al. 207). Zareipour et al. 2010 shows that an 1% improvement in the mean absolute percentage error (MAPE) would result in about 0.1–0.35% cost reductions from short term electricity price forecasting. Gaspain, Lukovic & Alippi (2019) provides some review on electricity prices methodologies and performances evaluation.

Despite the success of Deep Learning (DL) in all these energy-related areas and time series forecasting applications, there has not yet been, to the best of our knowledge, an attempt to bring its ideas and models to the field of electricity price forecasting. To make headway on this question, we implement and compare several Deep Learning algorithm in the electricity price forecasting

4 - Data sources

The data used to predict the DA and RT prices is public data provided by the PJM ISO. For this project, we have used hourly demand forecast data, gas price forecast data, and wind power forecast data for the PJM market and for the MISO market. We also added variables for the percentage change in demand forecast from one hour to the next. For example, if the load forecast for hour five is 3450 MW and the forecast for hour six is 3639, the variable "percentage change in demand" for hour six will be 5.48%. The target variables were the DA and RT prices for each hour for each transaction point from January 2018 to December 2019. Table x shows the details of the data used to construct the predictive models.

Table 1. Variables used to predict the DA and RT prices of PJM and their source

Variable	Source
Hourly DA price	PJM public data. Available for every trading point
Hourly RT price	PJM public data. Available for every trading point
Forecasted gas price	PJM public data
Demand forecast for PJM zones (Mid-Atlantic, Western and total)	PJM public data. Available for every trading point
Percentage of variation between demand forecast between two consecutive hours	Calculated from the demand forecast data.
Wind energy production forecast for PJM and MISO markets (MISO, MISO north, MISO central)	PJM public data. Available for every trading point

During the data collection and cleanup process, the missing data was imputed by replacing the missing data with the average of the last three available data points. We are conscious that this decision will smooth our data and introduce bias. We will take this into account during our analysis.

5 - Methodology

During this project, we used deep learning techniques to predict the DA and RT prices of the SMECO_HUB transaction point. We selected this point because it was the most profitable in the last year. Rather than just try a single Deep Learning method, we decided to try three types of Deep Learning models to compare effectiveness and further explore the application of Deep Learning to energy prices. The three methods used include multilayer perceptron network-based networks, convolutional neural networks, and recurrent neural networks.

5.1 - Multilayer Perceptron Networks

Generally, neural networks like Multilayer Perceptron or MLPs provide capabilities that are offered by few algorithms, such as:

(1) Robust to Noise: Neural networks are robust to noise in input data and in the mapping function and can even support learning and prediction in the presence of missing values.
(2) Nonlinear: Neural networks do not make strong assumptions about the mapping function and readily learn linear and nonlinear relationships.
(3) Multivariate Inputs: An arbitrary number of input features can be specified, providing direct support for multivariate forecasting.
(4) Multi-step Forecasts: An arbitrary number of output values can be specified, providing direct support for multi-step and even multivariate forecasting.

For the Day Ahead prices, the normalisation of all variables between 0 and 1 achieves the best performance. The best performing architecture for the Day Ahead prices is given below.

Table 2. DA architecture for MLP modeling

Parameters	Entrant layer	Hidden Layer	Output Layer
Number of layers	1	2	3
Number of neurons	13	5	3
Activation	relu	relu	linear
Number of epochs	1000		
Optimizer	Adam		
Regularization	« Early stopping »		

For Real time prices, the normalisations of the variable lead to very bad performance and the model doesn't converge too often; then the variables are used without normalisation/standardisation. After tuning, the retained architecture for the Real time prices is provided below.

Table 3. RT architecture for MLP modeling

Parameters	Entrant layer	Hidden Layer	Output Layer
Number of layers	1	1	3
Number of neurons	13	9	3
Activation	relu	relu	linear
Number of epochs	1000		
Optimizer	Adam		
Regularization	« Early stopping »		

5.2 - Convolutional Neural Network

Convolutional Neural Networks (CNN) is a Deep Learning algorithm traditionally used in the field of computer vision and image classification. CNNs are regularized versions of multilayer perceptron networks, taking advantage of the hierarchical patterns in data and assembling more complex patterns using smaller simpler patterns. Compared to other classification algorithms CNN requires significantly less pre-processing.

The process of building this model began with considering how many time steps one should use. In other words, how much of the past data should one use to predict future data values. In the case of energy prices, partial auto-correlation was analysed on both DAM and RT separately to determine how many time steps one should use for the CNN model. As can be seen in figure 1, the partial autocorrelation stays significant for the first 4 lags for DA and 25 lags for RT.

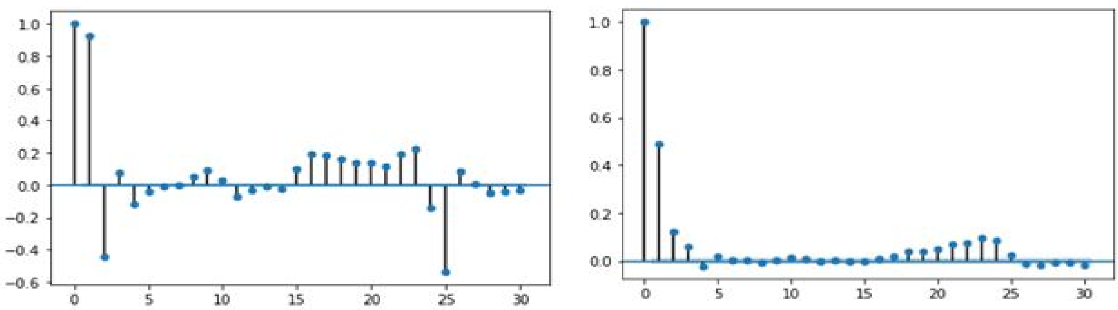


Figure 1 - Partial Autocorrelation analysis for DA (left) and RT (right) prices

Next, the data was normalized in order to improve stability and performance. Unscaled input variables can result in a slow or unstable learning process, whereas unscaled target variables on regression problems can result in exploding gradients causing the learning process to fail [Jason Brownlee]. Following this, the dataset was separated into a training set and test set to measure model performance.

Before building the actual model, the input matrix was created using the time steps from the partial autocorrelation analysis. The CNN model was developed using the Keras library, and used a relu activation function, mean squared error as the loss function, and ADAM as the optimizer. The same method was used on both DA and RT data.

5.3 - Recurrent Neural Network

For this project, we choose the Long Short -term memory architecture of recurrent neural networks. LSTMs are a type of gated recurrent neural network which incorporate the concept of cell state, which can be thought of as the memory of the network and that is updated at each time step by adding or removing information from it. The flow of information to and from the cell state is controlled by 3 gates: forget, input, output. Each gate is composed of a sigmoid neural network layer and a pointwise multiplication operation. The sigmoid layers output values between 0 and 1, controlling the amount of information that is let through the gates; 0 means nothing is let through, while 1 means that everything is let through. Just as the inputs at each time step are used to update the memory of the LSTM, the cell state serves as a filter that contributes to the definition of the output (Hochreiter and Schmidhuber 1997).

To use this kind of architecture, the data has to be reshaped into a three dimensional matrix. The first dimension is the batch size, the second is the number of steps in the past we want to consider and the third is the number of features. The batch size for both the DA and RT models is 24 to get the last 24 hours prior to the predicted price, the number of steps was decided according to the partial autocorrelation graphics (figure X): 25 for DA prices and 3 for RT prices.

The hyperparameters that we decided to tune are the number of neurons on the LSTM layer and the number of hidden dense layers. For the SMECO point, the best DA model had 90 neurons on the LSTM layer and one dense layer. The best RT model had 100 neurons on the LSTM layer and one hidden dense layer. As for the other hyperparameters, we chose widely used parameters: learning rate of 0.001, the activation function for the dense layers is reLu, for the output layer is sigmoid and for the LSTM layers is the default of Keras; tanh and sigmoid for the recurrent step. To avoid overfitting, we added early stopping with a minimum delta of 0.1 and a maximum of 30 epochs. To finish the model description, RMSprop was the optimizer and MSE was the loss function.

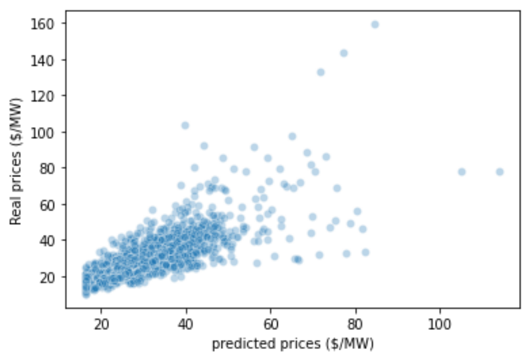
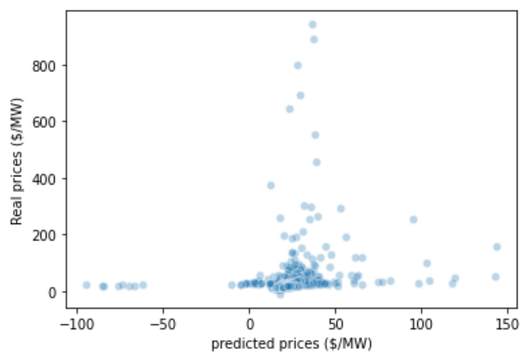
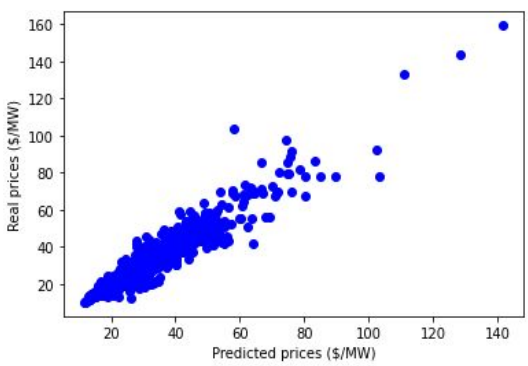
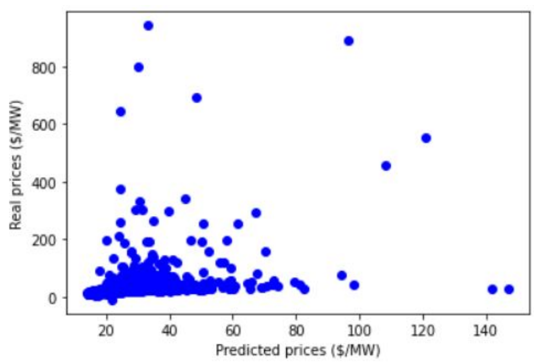
6 - Results and discussion

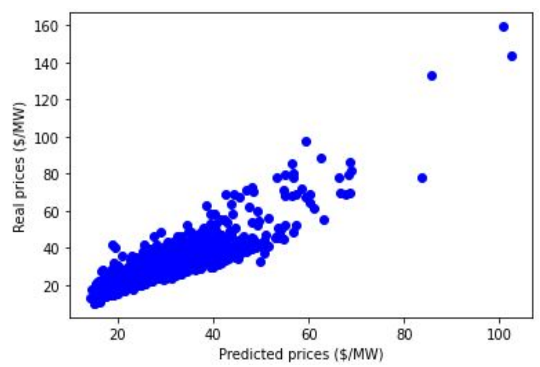
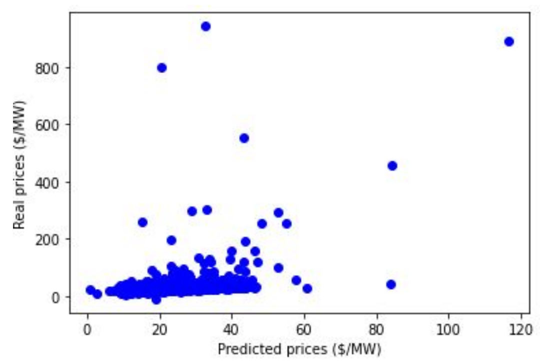
As can be seen in the table 4, different models obtained different accuracies. The table 5 shows the real price versus the predicted price for the DA and RT models. We observed that the DA prices were better predicted than RT prices. This is expected since we used the same features to predict both but, the RT prices are much more volatile than DA prices. To predict better RT prices, we hypothesize that additional information would be required about the network state, such as programmed equipment downtime. The MSE value for RNN is very high, which when observed in more detail revealed that for certain hours (9 instances in total) the model predicted higher prices than the real prices. Nevertheless, on average the model has an error between 5$ and 10$. The best overall model was found to be MLP.

Table 4. MAE and MSE on test set for the MLP, CNN and RNN methods

	Method	Day Ahead price	Real-Time price
MSE	MLP	7.09	35.37
	CNN	7.73	870.77
	RNN	118.81	1117.89
MAE	MLP	4.75	7.96
	CNN	9.04	7.32
	RNN	5.82	10.93

Table 5. Real prices ($\MW) versus predicted prices ($/MW) for DA (left) and RT (right) models

<table>
<tr><td></td><td>Day Ahead Price</td><td>Real- Time Price</td></tr>
<tr><td>MLP</td><td>
</td><td>
</td></tr>
<tr><td>CNN</td><td>
</td><td>
</td></tr>
</table>

RNN		

7- Conclusion

Having the ability to accurately predict Day Ahead and Real-Time electricity prices can enable businesses involved in the energy market to make more informed decisions and hence greater productivity and profit. Although the energy markets are very complex and involve many moving pieces, this report has shown very positive results in using Deep Learning models to tackle this challenge. This performance was achieved with relatively few demand variables and lag prices, which led to less input and computational burden of the curse of dimensionality. Although these methods can be implemented for real-time energy prices, this report found that Day Ahead signals tend to be more robust to noise and hence less likely to produce false positives when compared with Real Time prices. Overall, the best performing deep learning algorithm was found to be MLP and it is capable of identifying the changes in trends. To build upon the findings of this report, we suggest further research into the MLP model to incorporate more information and improve accuracy. Perhaps one can use economic makers such as unemployment rate, consumer price index, or industrial production to get a better sense of energy needs in the PJM market.

References

https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/

Jonathan Matherm, Eliyan Bitar, and Kameshwar Poolla. "Virtual Bidding: Equilibrium, Learning and the Wisdom of Crowds." IFAC Papers Online 50, no. 1 (2017): 225-32.

Sevi Baltaoglu, Lang Tong, and Quing Zhao. "Algorithmic Bidding for Virtual Trading in Electricity Markets." Journal of LaTex Class Files 14, no. 8 (August 2015).

Rafal Weron. "Electricity Price Forecasting: A Review of the State-of-the-art with a Look into the Future." International Journal of Forecasting, no. 30 (2014): 1030-081.

Aurélien Géron. Hands-On Machine Learning with Scikit-Learn & TensorFLow. 1st ed. Sebastopol, CA: O'Reilly Media, 2017.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org

Ian Pointer. Programming Pytorch for Deep Learning. 1st ed. Sebastopol, CA: O'Reilly Media, 2019.

Electricity price forecasting: a review of the state-of-the-art with a look into the future R. Weron Int J Forecast, 30 (4) (2014), pp. 1030-1081, 10.1016/j.ijforecast.2014.08.008

B. Szkuta, L. Sanabria and T. Dillon. Electricity price short-term forecasting using artificial neural networks. IEEE Trans Power Syst, 14 (3) (1999), pp. 851-85710.1109/59.780895

Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Proceedings of the 27th international conference on neural information processing systems. NIPS'14. MIT Press; 2014. p. 3104–12.

L. Jiang and G. Hu, "Day-Ahead Price Forecasting for Electricity Market using Long-Short Term Memory Recurrent Neural Network," 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, 2018, pp. 949-954.

Y. Zhu, R. Dai, G. Liu, Z. Wang and S. Lu, "Power Market Price Forecasting via Deep Learning," IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, 2018, pp. 4935-4939.

C. Feng, M. Cui, B.-M. Hodge, J. Zhang. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting, Appl Energy, 190 (2017), pp. 1245-1257, 10.1016/j.apenergy.2017.01.043

H.-Z. Wang, G.-Q. Li, G.-B. Wang, J.-C. Peng, H. Jiang, Y.-T. Liu, Deep learning based ensemble approach for probabilistic wind power forecasting, Appl Energy, 188 (2017), pp. 56-70

H. Zareipour, C. A. Canizares and K. Bhattacharya, "Economic Impact of Electricity Market Price Forecasting Errors: A Demand-Side Analysis," in IEEE Transactions on Power Systems, vol. 25, no. 1, pp. 254-262, Feb. 2010.

Alberto Gasparin, Slobodan Lukovic, Cesare Alippi. (2019). Deep Learning for Time Series Forecasting: The Electric Load Case. https://arxiv.org/abs/1907.09207