Warm-start Gradient Tree Boosting for Short-term Load Forecasting

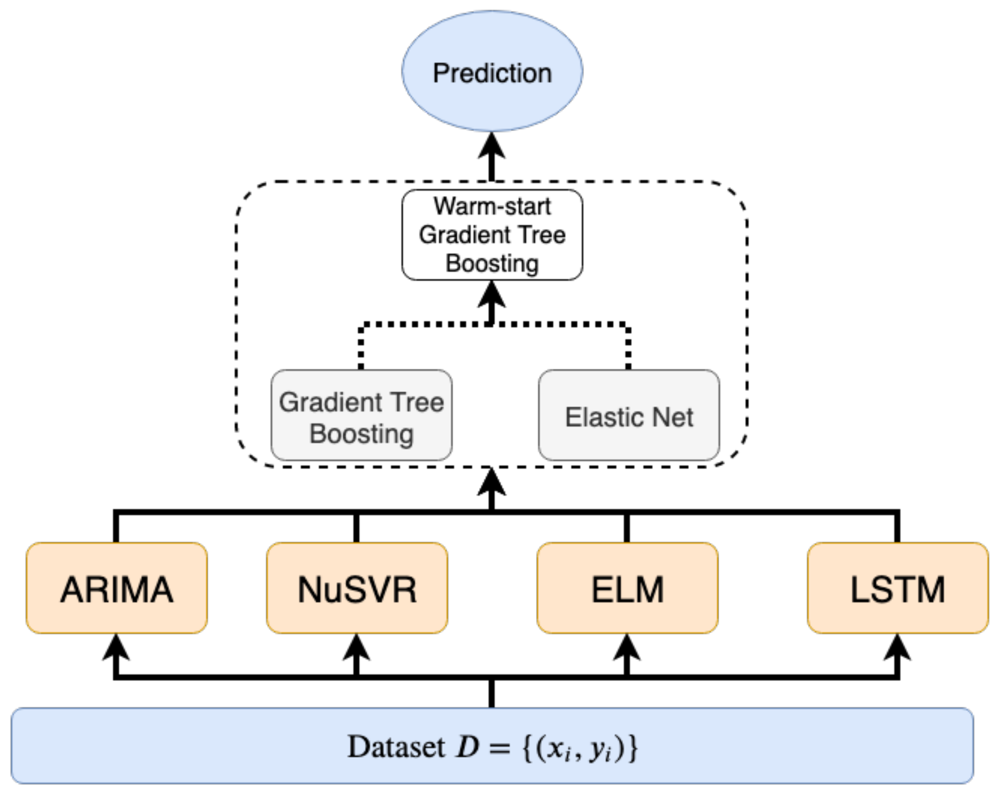
Jiahong Wang
Department of Electrical and Computer Engineering, Grainger College of Engineering, University of Illinois at Urbana-Champaign

INTRODUCTION

Electric power is a non-storable product, electrical operators should ensure a precise balance between the electricity production and consumption at any moment. Therefore, load forecasting plays a vital role in the daily operational management of power utility, such as energy transfer scheduling, unit commitment, load dispatch, and so on. With the emergence of load management strategies, it is highly desirable to develop accurate load forecasting models for these electric utilities to achieve the purposes of higher reliability and management efficiency.

METHOD

Overview
I propose an ensemble model for short-term load forecasting in the field of electric grid where electric power is a non-storable energy and needs precise forecast to minimize waste. Whereas traditional methods use a single linear or nonlinear forecasting algorithm, my ensemble model forecasts based on a combination of outputs from both linear and nonlinear algorithms. So my model gains higher accuracy by taking both linearity and non-linearity into account. Though conventional ensemble algorithms, such as Gradient Tree Boosting (GTB) and ElasticNet, perform decent jobs on my dataset, they merely take either linearity or non-linearity into consideration. So I propose Warm-start Gradient Tree Boosting (WGTB) that hybrids ElasticNet and GTB and outperforms existing ensemble algorithms by taking both linearity and non-linearity into account.



METHOD

ARIMA
Autoregressive Integrated Moving Average (ARIMA) is a generalization of Autoregressive Moving Average (ARMA) that integrates Autoregressive (AR) and Moving Average (MA) into a combined time-series prediction model. It is often written as ARIMA (p, d, q). Its parameters p, d, q respectively denotes number of autoregressive terms, lag order of non-seasonal difference and number of moving average terms.

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t + \sum_{i=0}^{q} \theta_i \epsilon_{t-i}$$

NuSVR
The Nu Support Vector Regression (NuSVR) proposed by Scholkopf et al has the advantage of using a parameter Nu for controlling the number of support vectors. The regularization parameter C in the ordinary SVR formulation is replaced by a parameter Nu which is bounded by 0 and 1.

METHOD

ELM
Extreme Learning Machine (ELM) is a special case of single-hidden-layer fully connected neural network. Unlike neural network where weights of both two layers are trainable, only the weights of output layer is trainable and the weights of the hidden layer is randomly initialized and immutable during training. This benefits ELM to have a global optimum.

LSTM
Recurrent neural network (RNN) is designed for time series data forecasting; therefore it is a good fit for load forecasting. Unlike traditional feed forward neural network where inputs of different timestamps are fed into the network together, the inputs of RNN are fed into the network sequentially. However, RNN model suffers from gradient vanishing problem. This prevents weights from changing value and, in some cases, it stops the model from further learning. To solve this issue, Hochreiter and Schmidhuber introduced a Long Short-term Memory (LSTM) that includes an extra memory cell.

METHOD

ElasticNet
ElasticNet is a generalized version of linear regression. It includes both L1 and L2 regularization terms of the coefficients, which allows it to learn a sparse model like Lasso while maintaining regularization properties of Ridge.

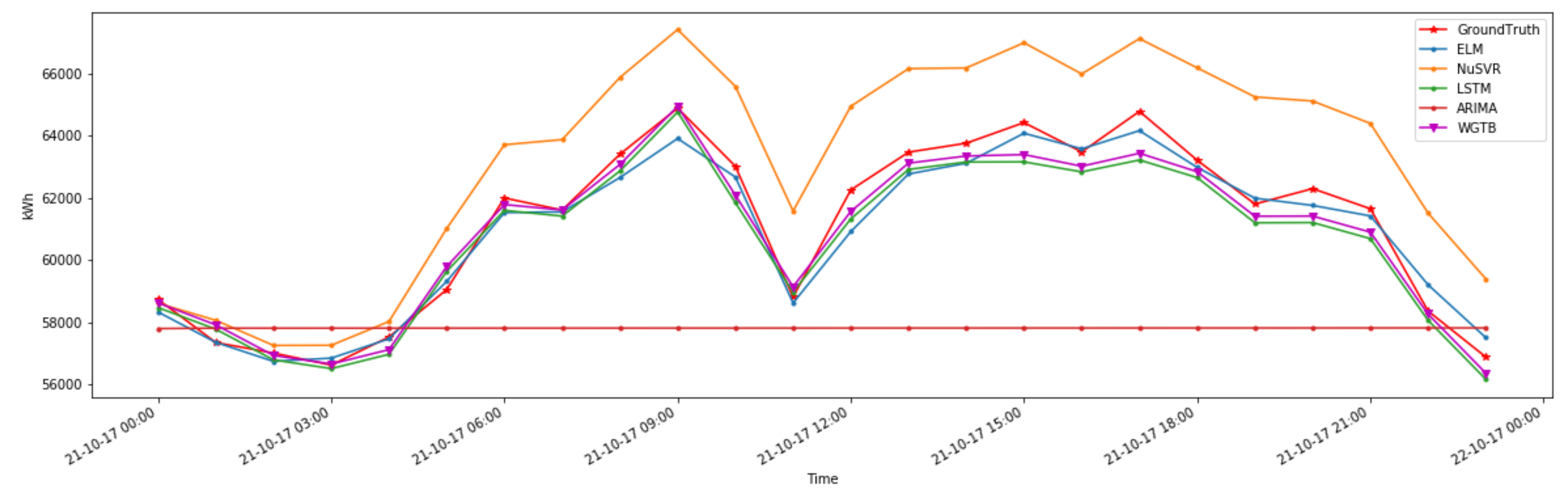
Gradient Tree Boosting
Gradient Tree Boosting (GTB) was first introduced by Friedman. The model is a generalized greedy function approximator using boosting method. Each weak regressor in the model is a decision tree, namely CART (Classification and Regression Trees). The mathematical formulation is in additive fashion.

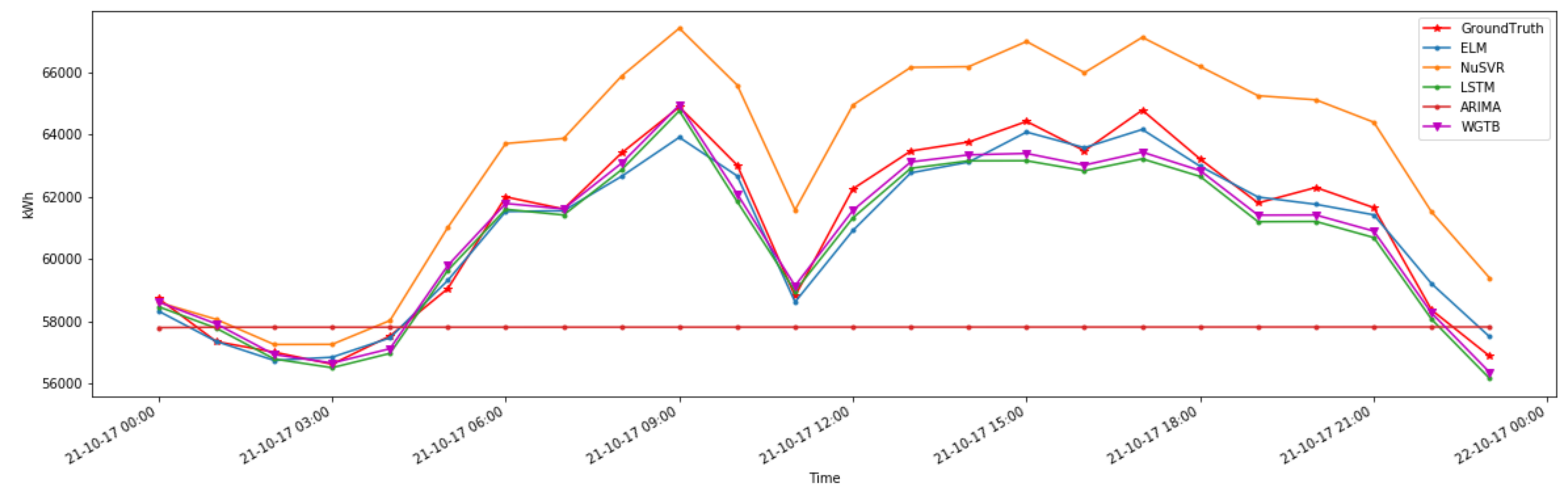
$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \nu \gamma_m h_m(\mathbf{x})$$

$$h_m(\mathbf{x}) \leftarrow -\nabla_F \sum_{i=0}^{N-1} L(y_i, F_{m-1}(\mathbf{x}_i))$$

RESULTS

The two figures below visualize and compare the predictions from ensemble model versus single submodels in two arbitrary samples. The samples are drawn from hourly electric load consumption data of Jiangsu Province in China. The figures strongly reflect the improved accuracy of the proposed Warm-start Gradient Tree Boosting (WGTB).





CONCLUSIONS

I have proposed WGTB, an new ensemble forecasting model that outperforms existing models by taking both linearity and non-linearity into account in two ways. First, outputs of the proposed model are forecasted based on the juxtaposition of outputs from both linear and nonlinear submodels. Second, during ensemble process, the proposed model hybrids linear ensemble model and nonlinear tree ensemble model. Experiments have been conducted to prove the effectiveness of our proposed model in the improvement of accuracy and robustness.

ACKNOWLEDGEMENTS

I would like to express my special thanks to Ms. Zhang who patiently instructed me in data collection, methodology development, data analysis and manuscript composition.

NUS Engineering
National University of Singapore

ILLINOIS