Short-Term System Marginal Price Forecasting Using 2-Stage Variable Selection Approach

[SangWoo Shim, Konkuk University, +82-10-2368-0294, tkddn349@konkuk.ac.kr]
[DaHan Lee, Konkuk University, +82-2-458-4778, timeflaw@konkuk.ac.kr]
[JaeHyung Roh, Konkuk University, +82-2-450-3934, jhroh@konkuk.ac.kr]
[Jong-Bae Park, Konkuk University, +82-2-450-3483, jbaepark@konkuk.ac.kr]

Overview

The Korean electricity market structure is a CBP(Cost Based Pool). Due to this type of electricity market, the cost of a Genete unit is set by KPX(Korea Power Exchange). Generation companies have to bid only for capacity when they participate in the electricity market. Therefore, information on SMP(System Marginal Price) may seem unnecessary. Still, SMP is an important component of the electricity market[1]. Suppose the Korean electricity market changes to the PBP(Price-Based Pool) market. In that case, forecasted SMP will contribute to establishing strategies for price bidding by generators.

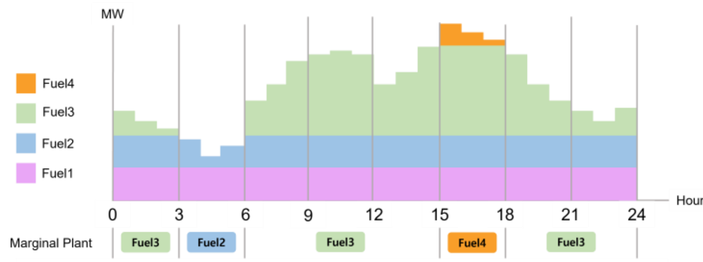

Figure 1. SMP determination

KPX(Korea Power Exchange) forecasts the load and receives offers for capacity from generation companies one day ahead. It then determines the market price by producing a PSS(Price Setting Schedule) set up by a program that minimizes the total production cost of generating units. In the PSS, SMP for each hour is calculated to meet the demand for each hour. The marginal price of the most expensive generating unit in the PSS is the SMP, and that unit becomes the marginal plant(the last plant in merit order)[2].

Many studies in forecasting SMP have included predictions using rough set theory and ANN(Artificial Neural Network) and long-term forecasting using a periodic pattern of SMP decomposition and oil price[3-4]. This paper suggests a 2-Stage method that predicts SMP at the second stage using load from the first stage. The electricity demand data is important to forecast SMP. Therefore, predicting SMP through load forecasting is helpful for operators with sparsity information.

To forecast SMP, models are used, such as Support Vector Machine(SVM), Random Forest, XGBoost(eXtream Gradient Boost), and LSTM(Long Short-Term Memory). Because they are still often used nowadays and have each strength. SVM model has less impact on error data, XGBoost and Random Forest do not overfit well, and LSTM is proper for time series forecasting. The algorithm trains data from the first in January to the before test day using forecasting models and calculates predicted SMPs. The prediction performance based on the ensemble of the above models is expressed through MAPE(Mean Absolute Percentage Error), nRMSE(normalized Mean Absolute Error), nMAE(normalized Mean Absolute Error) in the Result sector, and future research plans to develop the paper are written in the Conclusions.

Methods

Predicting algorithm for SMPs using the 2-Stage method is illustrated in Figure 2. At the load forecasting stage, weather data, including temperature and amount of sunlight, historical demand of a day before, and time information(i.e., hour, month, and day of the week) are used. Weather data provided by the Korea Meteorological Administration[5] are classified by region. Therefore, in this paper, national weather data were constructed by weighting based on the amount of regional load consumed[6]. One of the above forecasting models is used for predicting load and the last three forecasting models repeat the same process.

Before load predicting, important variables need to be selected to increase the performance. The algorithm uses SHAP(Shapley Additional exPlanations) method and the Pearson Correlation Coefficient method. The SHAP method measures the importance of variables, and the Pearson Correlation Coefficient method measures linearity between variables.

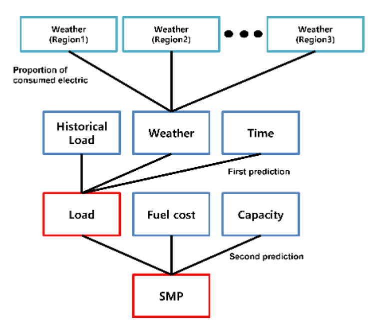

Figure 2. 2-Stage method

The variables with high linearity in prediction need not be used together and decrease forecasting performance, so this paper removed them.

The SHAP method is an artificial intelligence method that explains the relationship between the input variable and the result value of the forecasting model through Shapley Value[7]. The importance of variables for forecasting load is expressed in Figure 3. The Shapley Value is obtained according to the change due to the addition or removal of the corresponding variable from the combination of various variables.

The Pearson Correlation Coefficient indicates the degree of linear relevance between two variables and has a value between –1 and 1. The more significant the absolute value of the Pearson correlation coefficient, the more significant the correlation, and the sign indicates the direction of the correlation. When the Pearson correlation coefficient is expressed in Figure 4, it confirms that all variables have a low correlation, except that the correlation between the month variable and the temperature variable is slightly higher at 0.67.

At the SMP forecasting stage, SMPs are calculated using load forecasting results, fuel costs of each generator such as nuclear, coal, natural gas, oil, and the capacity of each fuel resource. This paper used the algorithm to conduct SMP predictions quarterly for the season in 2021, and testing was born three days after the training. All models were applied to each stage prediction. Thus experiments were conducted with a total of sixteen ensemble models.

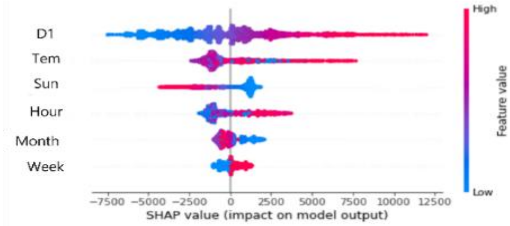

Figure 3. SHAP value of the input variable

Figure 4. Pearson Correlation Coefficient for input variables

Results

The XGBoost-Random Forest ensemble method makes the best performance for forecasting SMP. When forecasting SMP, the Random Forest model's case using load data from XGBoost performs best. That ensemble method performed about 1% better in MAPE than other ensembles, and compared to other studies using Artificial Neural networks, the performance was measured slightly better. And in general, XGBoost performs the best for forecasting load, and Random Forest performs the best for forecasting SMP. In demand forecasting, six variables were selected through the SHAP method for sixteen weather, time, and historical demand variables. The prediction performance was increased by about 0.5% in MAPE by removing unnecessary variables.

Conclusions

After classifying dates such as weekends, weekdays, and special days according to load patterns, the forecasting performance will be measured, and the recently presented forecasting model will be used for that algorithm. Suppose these algorithms include long-term forecastings such as demand, fuel cost, and capacity. In that case, it will be possible to establish a long-term SMP outlook, contributing to the generator installation, capacity planning, or investment.

References

[1] J. H. Kong et al. (2018), Development of ESS Scheduling Algorithm to Maximize the Potential Profitability of PV Generation Supplier in South Korea, Journal of Electrical Engineering & Technology, 13(6), 2227-2235.

[2] Korea Power Exchange, "https://new.kpx.or.kr/menu.es?mid=a10401040000".

[3] J. K. Lee, J. B. Park, J. R. Shin, & K. Y. Lee (2005), A system marginal price forecasting based on an artificial neural network adapted with rough set theory, IEEE Power Engineering Society General Meeting, vol. 1, 528-533.

[4] B. R. Oh, D. -H. Lee, & D. H. Lee (2022), Oil-Price Based Long-Term Hourly System Marginal Electricity Price Scenario Generation, in IEEE Access, vol. 10, 25051-25061.

[5] Korea Meteorological Administration, "http://www.weather.go.kr/.".

[6] D. H. Kim, H. J. Jo, M. S. Kim, J. H. Roh, & J. B. Park (2019), Short-term load forecasting based on deep learning model, The Transactions of The Korean Institute of Electrical Engineers, 68(9), 1094-1099.

[7] Eyal Winter (2002), Handbook of Game Theory with Economic Applications, vol. 3, 2025-2054.