Hourly Steel Industry Energy Consumption Prediction Using Machine Learning Algorithms

Sathishkumar V E, Myeong-Bae Lee, Jong-Hyun Lim, Chang-Sun Shin, Chang-Woo Park, Yong Yun Cho
Dept. of Information and Communication Engineering, Sunchon National University

Abstract

Predictions of Energy Consumption for Industries gain an important place in energy management and control system, as there are dynamic and seasonal changes in the demand and supply of energy. This paper presents and discusses the predictive models for energy consumption of the steel industry. Data used includes lagging and leading current reactive power, lagging and leading current power factor, carbon dioxide (tCO2) emission and load type. In the test set, four statistical models are trained and evaluated: (a) Linear regression (LR), (b) Support Vector Machine with radial kernel (SVM RBF), (c) Gradient Boosting Machine (GBM), (d) random forest (RF). Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used to measure the prediction efficiency of regression designs. When using all the predictors, the best model RF can provide RMSE value 7.33 in the test set

Keywords—Energy Consumption, Machine Learning, Random Forest, Linear Regression, Gradient Boosting Machine, Support Vector Machine.

1. Introduction

South Korea's production industry has begun to evolve at an elevated pace since the 1990s and has become the primary pushing power behind South Korea's continued fast economic development. In the 1990s, primary power usage grew at an annualized pace of 7.5%, which in the same era was greater than the annualized financial development level of 6.5%. This is due to strong development in energy-intensive sectors, including petrochemical sectors. The strong increase in industrial electricity consumption helped to boost the reduction of energy conversion, which further subverted energy intensity. The increase in energy production after 2009 significantly buffered the country against the global financial crisis but adversely affected the overall energy efficiency of the country [1]. Many uncertain variables, such as industrial structure, technology level, energy cost, financial scale and national policy, affect the energy consumption of the industries.

With the growing issue of coal and oil shortages, the energy resource-related forecast and planning issues have drawn considerable scrutiny from both the research and industrial practice perspective. To make reasonable use of by-product gasses in the steel industry, scheduling operators must be aware of the quantities of generation, consumption and storage overtime in real-time. Thus, the precise prediction of these energy flow units will provide a good guide to their planning and distribution. The iron and steel industries are always energy-intensive, covering 10% of the full industry's energy consumption. Recently, with the rising energy resource shortage, the energy supply condition in the Iron & Steel sector has become highly challenging. Developing an energy-saving policy has become an increasingly prominent job that can be achieved in respects such as technical advancement, refurbishment of machinery and enhancement of management. With rising energy prices, energy consumption costs are 10-20 times higher than that of the entire production of Iron & Steel. High energy usage will certainly lead to higher costs for Iron & Steel products and will result in more pollution and emissions. To this purpose, certain steps, such as optimizing the manufacturing framework and speeding up the development and promotion of techniques to save energy and reduce emissions, are required to guarantee the efficient supply of energy in the manufacturing industry in South Korea.

2. Related Works

A long-term prediction of energy consumption is proposed in [2] using a granular computing method that incorporates industrial-driven semantics and granulates initial data depending on the specificity of manufacturing processes. To quantify the performance of the proposed method, the authors used real-world industrial energy data from a steel plant in China. The experimental results show that the proposed method is superior to some other data-driven methods and can meet the requirements of the practically viable prediction.

Support vector machine (SVM) classifier is designed to predict the energy consumption level of the Ironmaking process in [3]. Particle swarm optimization (PSO) is introduced to improve accuracy for optimizing SVM parameters. First, the Ironmaking process's consuming structure is analyzed to accurately model the prediction problem. Then they presented the improved SVM algorithm. Finally, the experimental test is carried out based on a Chinese Iron and Steel company's practical data. The results show that the proposed method can predict with satisfactory accuracy the consumption of the Ironmaking process addressed.

A homologous grey prediction model with one variable and one first-order equation (HGEM (1,1)) is proposed to forecast Chinese production's total energy consumption based on the Grey system theory [4]. Using this model, they forecast China's manufacturing industry's complete energy utilization over the years 2018–2024. The findings indicate that Chinese manufacturing's complete energy consumption slows down but is still too massive.

Energy consumption and greenhouse emission forecasting for India's pig iron manufacturing organization are addressed in [5], as executives are concerned in knowing the present and future trends of these indicators for smarter environmental policy. Autoregressive Integrated Moving Average (ARIMA) is used for forecasting purposes to reveal that ARIMA (1,0,0) x (0,1,1) is the best energy consumption model. ARIMA (0,1,4) (0,1,1) is the finest equipped model for Greenhouse emissions. In both cases, the forecasts are similar to those of the seasonal random trend model but appear smoother as the seasonal pattern and trend are efficiently averaged for both energy consumption and GHG emissions.

A generic model for the power-consuming device specification is discussed in [6]. A tree-based compositional approach promotes arbitrary levels based on the machine's structure or external variables such as company policies. As the designs are placed in ontologies, this strategy is extremely extensible. Secondly, for each structural level, a methodology is suggested for static and dynamic modelling of power consumption. The prediction can be done based on that model. Furthermore, an instance is given for implementing and predicting a continuous casting machine process.

3. Data Description

The data is gathered from DAEWOO steel.co. Ltd in Gwanyang, South Korea. The Industry produces several kinds of coil, steel sheets and iron plates. The data on energy consumption is filed on the website of the Korea Electric Power Corporation (pccs.kepco.go.kr).

This analysis focuses on the Energy usage(kWh) information recorded for the industry every 1 hour. The data period is 365 days (2018, 12 months). Table 1 provides the load type and timing of each month. Fig 1 Displays the profile of energy consumption over the interval and elevated variation. Since the steel industry in open space and has no heaters or cooling facilities, the temperature variables have no impact on energy consumption. The overview of the full dataset is shown in Table 2. Other additional features are produced from the date/time variable: the number of seconds from midnight for each day (NSM), week status (weekend or weekday), and the day of the week.

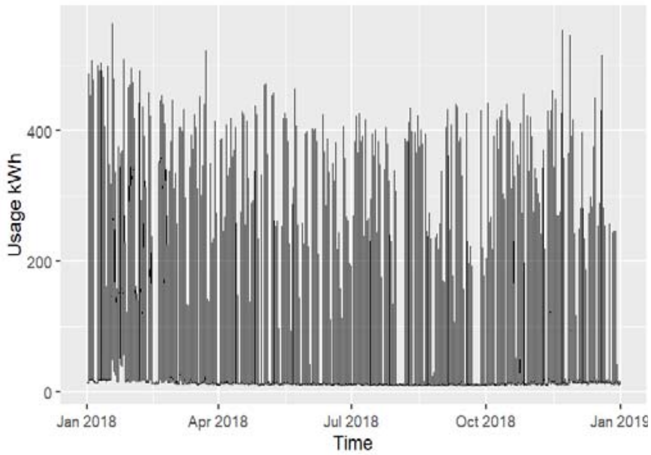
<Table 1> Load Type and Its Timings

Load Type	June-August	March-May,September-October	November - February
Light Load	23:00-09:00	23:00-09:00	23:00-09:00
Medium Load	09:00-10:00 12:00-01:00 17:00-23:00	09:00-10:00 12:00-01:00 17:00-23:00	09:00-10:00 12:00-07:00 20:00-22:00
Maximum Load	10:00-12:00 01:00-17:00	10:00-12:00 01:00-17:00	10:00-12:00 17:00-20:00 22:00-23:00

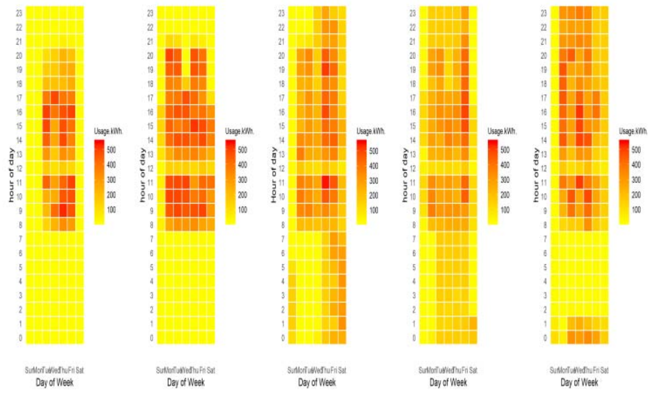
For five consecutive weeks, an hourly heat map is produced to identify any time trends and shown in Fig 2. It shows that the energy consumption pattern of the steel industry has a powerful time component. Energy consumption is lower during the weekend than the other days.Energy consumption begins to rise from 8 a.m. and maintains high up to 8 p.m.

<Table 2> Data Variables and Description

Data Variables	**Abbreviation**	**Type**	**Measurement**
Industry Energy Consumption	Usage	Continuous	KWh
Hour of the Day	Hour	Continuous	Hour
Lagging Current reactive power	LagRP	Continuous	KVarh
Leading Current reactive power	LeadRP	Continuous	KVarh
Lagging Current power factor	LagPF	Continuous	%
Leading Current Power factor	LeadPF	Continuous	%
tCO2(CO2)	CO2	Continuous	Ppm
Week status	Wstatus	Categorical	(Weekend (0) or a Weekday (1))
Day of week (Monday Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday)	Dweek	Categorical	Sunday, Monday …. Saturday
Load Type	Ltype	Categorical	Light Load, Medium Load, Maximum Load



(Fig 1) Steel Industry Energy Consumption Measurement for 1 year



(Fig 2) First Five weeks Hourly Steel Industry Energy Consumption Heatmap

4. Model Selection

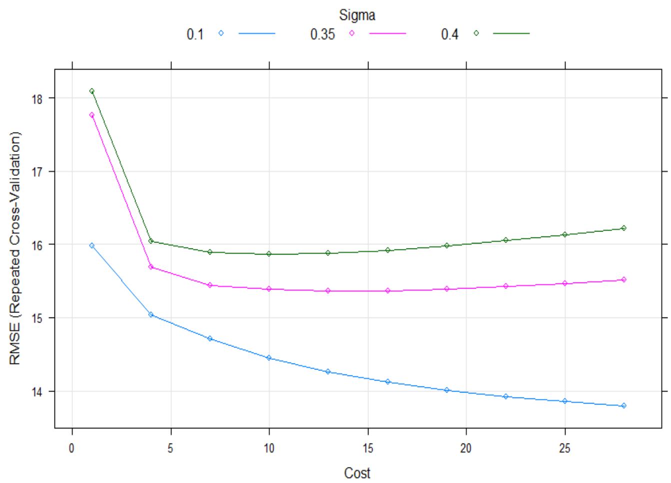
The complete one-year data set is split into training and test validation. 75% of the information is used for model training and 25% for testing purposes. The figures are indicated in Table 3

<Table 3> Training and Testing Set

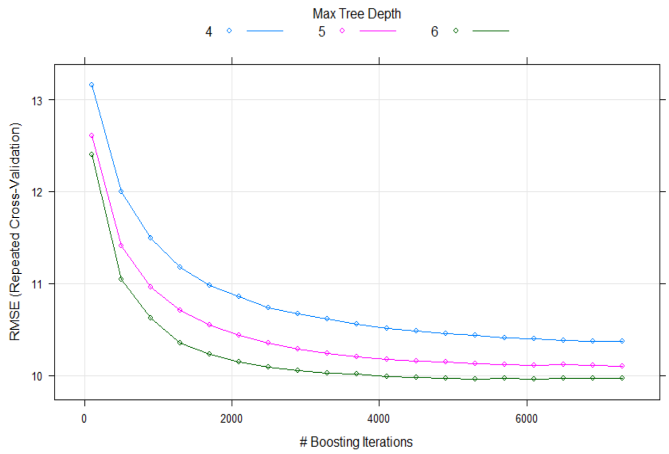
Dataset	Number of observations
Training	6572 and 10 variables
Testing	2188 and 10 variables

It is essential to find optimal tuning parameters for each of the regression algorithms for finding and reducing error values while designing a model. The outcomes of the grid search for SVM, GBM, and RF are presented in Fig 3, Fig 4, Fig 5 respectively.

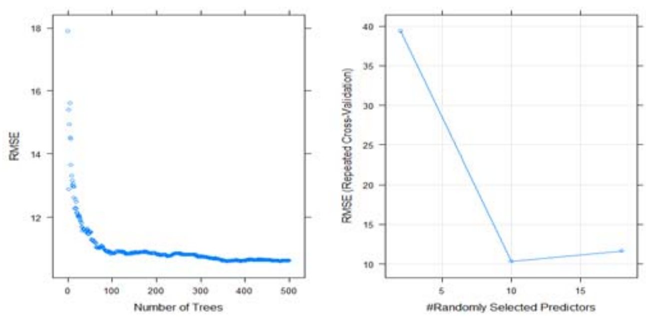
As indicated in Fig. 3, The optimal Sigma and Cost values for SVM RBF are 0.1 and 25 respectively. The optimal value for the number of trees for GBM is 5300 and the maximum depth of the tree is 6 as shown in Fig 4. In Fig 5, the RMSE value stays constant for RF after 400 and the randomly chosen predictors or mtry value is 10.



(Fig 3) Grid search results for optimal values of sigma and cost values for SVM-radial model



(Fig 4) Grid search results for optimal number of trees (Boosting Iterations and Maximum tree depth for GBM model



(Fig 5) RF with all the parameters

Table 4 shows the performance of each of the models. RF performance is best in the testing set yielding the lowest RMSE, MAE and MAPE values compared to other regression models. GBM's performance is also very close to the RF model and GBM's performance is even stronger than RF in the training set. Though GBM performance results are close to RF model, RF model is regarded to be the most precise model in this study.

<Table 4> Model Performance

Models	**Training**			**Testing**		
	RMSE	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**
LM	16.98	6.98	1.34	9.31	6.12	13.36
SVM	11.09	7.32	2.55	10.66	7.88	27.69
GBM	2.70	1.94	0.75	7.47	4.68	11.57
RF	5.12	2.57	0.63	7.33	4.60	9.89

5. Conclusion

One of the most important issues for energy management and optimization in the steel industry is a precise long-term prediction of energy consumption. The data analysis shows thought-provoking outcomes in the exploratory analysis. The purpose of this research is to determine the best performing machine learning techniques to predict the hourly energy consumption in the steel industry. The findings indicate that the RF model improves RMSE, MAE, and MAPE of predictions in consideration to other regression models considered in this research.

Acknowledgement

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20172010000730).

References

[1] Lee, Seung-moon, Mid-term Korea Energy Demand Outlook, Korea Energy Economics Institute, May 2014.

[2] Zhao, Jun, et al. "Granular model of long-term prediction for energy system in steel industry." *IEEE transactions on cybernetics* 46.2 (2015): 388-400.

[3] Zhang, Yanyan, Xiaolei Zhang, and Lixin Tang. "Energy consumption prediction in ironmaking process using hybrid algorithm of SVM and PSO." *International Symposium on Neural Networks*. Springer, Berlin, Heidelberg, 2012.

[4] Zeng, Bo, Meng Zhou, and Jun Zhang. "Forecasting the energy consumption of China’s manufacturing using a homologous grey prediction model." *Sustainability* 9.11 (2017): 1975.

[5] Sen, Parag, Mousumi Roy, and Parimal Pal. "Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization." *Energy* 116 (2016): 1031-1038.

[6] Reimann, Jan, et al. "Methodology and model for predicting energy consumption in manufacturing at multiple scales." *Procedia Manufacturing* 21 (2018): 694-701.