Application of machine learning methods for optimizing the technical and economic performance of generating systems

E K Arakelyan¹, I A Boldyrev², K V Evseev¹² and Y A Gorban¹²

¹National Research University Moscow Power Engineering Institute, Moscow, 111250, Krasnokazarmenaya str., 14, Russia
²Volzhsky branch of National Research University Moscow Power Engineering Institute, Volzhsky, 404110, Lenin Avenue, 69, Russia

Abstract. The paper is on the technical and economic performance optimization technique for thermal and power generating system using machine learning methods. The possibility of using regression analysis for parameter influence evaluation when calculating technical and economic performance in order to reach better generating unit efficiency is described. The approach to evaluate the parameter influence of a large distributed control system is presented.

1. Introduction

When producing power energy, it’s necessary to assess and take into account the generating system performance. In this paper, the particular case of a thermal power plant generating unit, which includes such components as steam boiler, turbine, condenser, pumps, etc., is concerned [1]. During power equipment operation, assurance of its efficiency, maintenance schedules and commissioning and running tests are crucial. This information can be obtained by calculating technical and economic efficiency indices (TEI). Most commonly the data to calculate TEI is transferred to the distributed control system (DCS) via information and measuring channels (IMC) from field devices and sensors, that usually represent a lot of parameters and values. Though its number can be great, it does not always lead to more precise results, but it often tends to complicate the assessments [2].

The number of parameters and signals in modern generating units can be roughly at least 1000, however not all of them can affect the result the same way, which means that it is possible to select the most important ones, which affect the result of TEI assessments and power equipment monitoring and control most, among all. It means that the more precise and actual these values, the more precise the DCS monitoring and control, thus the more efficient the generating system [3].

The goal of the research is a development of the technique for generating system influential process parameter selection, thus it is possible to take into account, which ones are the most important for an information and control system efficiency. The problem can be solved with the machine learning methods to select the parameters among the set of input values. The historical and real-time data can be used as training data for a system state prognosis model. The feature of machine learning is that it is capable to solve classification and regression analysis tasks on data that are not exactly linearly separable [3]. Using such a technique as machine learning is also useful for calculating the operational technical and economic performance of a thermal power plant generating unit for which 10-15 minutes prognosis is necessary for ensuring high efficiency [2].
2. Machine learning methods

Machine learning (ML) is a tool for solving a wide range of applied problems (pattern recognition, equipment diagnostic, state analysis and prognosis, etc.) based on mathematical statistics, optimization methods, probability theory, graphs theory and other mathematical techniques.

Figure 1. The general ML model flowchart.

The feature of ML is an indirect approach of getting the result of calculations, it requires building a certain model with a specific architecture and algorithms. After that step it should be trained, i.e. its properties (parameters) should be adjusted for solving the specific problem. Training is the main process in ML and it means finding the optimal parameters to maximize the output value, the result to correct answer based on dataset values [3]. In other words, the model should generalize the obtained data to calculate the result for new data that is not in the dataset. Using of any model can be considered a function (figure 1), which takes certain values as input and returns output values. Depending on the desired output it can be a continuous value (regression) or discrete (classification, clusterization) (figure 2) [3].

Figure 2. ML models. a - classification, b - clusterization, c - regression.
There are a few ML training approaches - supervised learning and unsupervised learning (figure 3). The former one requires both input (X) and output (Y) value (ground truth) sets when the last one generalizes the data by only vector X. In general, all ML approaches require dataset for training, another dataset can be used for the testing model, its accuracy.

Therefore, ML allows finding dependencies between values based on experimental data, however, it should be mentioned that getting an absolutely precise model is nearly impossible in most cases for there are always outliers in data values and during training, the training algorithm tries to find the parameters that minimize the error (the difference between ground truth Y and predicted output Ŷ) [3]. In other words, it requires finding the loss function global minimum to highest possible accuracy, which could be a challenge when the number of input features is great. Nevertheless, ML is a powerful tool in data analysis.

3. Use of ML methods for optimizing technical and economic performance
During the operation of thermal power plant generating units, it is necessary to ensure their efficiency, how well it transforms primary fuel to energy. To measure this value there are algorithms that take as input process parameters. It should be mentioned that distributed control systems operate a huge amount of values, however not necessarily all of them affect the generating unit performance equally, so there are parameters that are more important to monitor and control in order to reach higher efficiency. Thus it could be useful to select them and place greater focus on them what could be important for performance prediction within a certain range of time.
If the parameter values and results are given it is possible to analyze the data and find dependencies between them, how the former ones affect the result with regression analysis. One of the main purposes of regression analysis is determining the influence quantity of input values (predictors) on output ones and result from prediction under certain conditions [4].

The mentioned technique can be used for parameter selection from the great number of process variables, the output value is represented by technical and economic performance estimation of heat and power equipment. Nevertheless, in the case when there is a lot of input parameters it could be inconvenient to carry out regression analysis on all the parameters at the same time, thus it is reasonable to split the one complex task into a set of small ones with fewer input features and solve it separately. Such an approach could be represented as an oriented weighted graph, where each start node is an input parameter, end node is the result, performance estimation, other ones are intermediate outcomes (figure 4). The graph weights are determined with regression analysis and represent parameter influence. The sum of all incoming weights of a node equals 1, so first it should be normalized before the next step. Thus it shows the relative influence between parameters, the highest weight has the biggest influence on the node value.

![Figure 4. Parameter influence graph.](image)

The input parameter influence on the result can be found as a product of weights during the graph traversal [5]. The result of the procedure is the vector of each input parameter influence on output value (figure 5). In cases when one parameter is used in calculations more than once, the highest influence is recorded. The results allow concluding which parameters monitoring and control are more important for it leads to higher generating unit efficiency. Thus it could be necessary to improve the DCS measuring channels in order to provide it more data considering these parameters. This problem is a part of the intelligent distributed control system design, which means more efficient monitoring and control by data flow optimization and unifying multiple DCS levels within one framework, that allows considering an object performance overall [6, 7].

4. Conclusions

Application of machine learning methods for technical and economic optimization of generating unit allows analyzing and evaluate distributed control system process parameters in order to select the most influential ones, thus it is possible to increase generating unit efficiency by improving the information content of equipment operation;

The proposed approach of splitting one complex calculation task into many simple for parameter influence estimation is reasonable in the case when it is necessary to operate with a great number of input parameters.
References

[1] Pletnev G P 2005 Automation of Technological Processes and Production in Heat Power Engineering (Moscow: MPEI)

[2] Arakeljan E K and Galustjan M K 2012 Methodological approaches to the calculation of annual technical and economic indicators of power equipment of an autonomous complex for the production of electricity and heat New in Russian energy sector 5 16-28

[3] Raschka S and Mirjalili V 2019 Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-learn and TensorFlow 2, 3rd Edition (Birmingham: Packt)

[4] Draper N R and Smith H 1998 Applied Regression Analysis, Third Edition (New York: John Wiley & Sons, Inc.)

[5] Diestel R 2017 Graph Theory (New York: Springer)

[6] Arakeljan E K, Vasil’ev E D and Khurshudjan S R 2014 Automated control system issue and a possible way of its decision Vestnik MPEI 1 15-20

[7] Evseev K V and Arakelian E K 2020 Information support of the intellectual ACS TP Radio engineering, electrical engineering and power engineering (Moscow: Moscow Power Engineering Institute) p 919

Acknowledgments

The research is funded by RNF grant № 19-19-00601.