The technique for increase of generating system technical and economic indexes evaluation accuracy using machine learning

E K Arakelyan¹, I A Boldyrev², K V Evseev¹,², M M Sultanov² and V A Yurov²

¹National Research University "MPEI", Russia, 111250 Moscow, Krasnokazarmennaya, 14
²Volzhsky branch of National Research University MPEI, Volzhsky, 404110, Lenin Avenue, 69, Russia

kirillevseyev@gmail.com

Abstract. This paper addresses the problem of thermal power plant generating unit efficiency evaluation accuracy increase. The comparison of classic regression analysis and artificial neural network regression approaches is described. The feasibility and reasonability of application of the machine learning techniques for generating unit performance prediction and selecting most influential parameters on its efficiency evaluation accuracy are considered.

1. Introduction
The efficiency and operating economy of power equipment are critical criterions for generating system performance, in particular thermal plant generating unit, producing electrical and heat energy, which consists of such components as boiler, turbine, condenser, pumps and other equipment. Technical and economic indexes (TEI) evaluation gives information for ensuring most efficient equipment operation, repairs prediction, operating personnel performance evaluation and pre-commissioning activities. Operating TEI are evaluated as a rule at constant regular interval and it is necessary to consider change in the input parameters for the base period [1].

To perform technological parameter calculations and estimations there are computational stations – specific software application to arrange parameter and indexes computation using formulas and calculate their values by real-time or historical data [2]. Modern computational stations provide a wide range of tools to develop various calculation algorithms, nevertheless one of disadvantages of them is the limited prediction functionality [3]. Data required to evaluate TEI is acquired from sensors via informational channels, however the number of signals in distributed control system (DCS) can be great, what may complicate calculations and not necessarily increase the TEI evaluation accuracy. Among these parameters it is possible to select the parameters, which influence on the TEI result most, thus increase in monitoring them allows achieving more actual values.

The subject of the research is developing the technique for increase of generating system TEI evaluation accuracy based on the equipment operating parameters analysis. During heat and power equipment operation of generating system it is necessary to ensure high efficiency and operating economy. In DCS it is important to generate control actions before the event, that can lead to sufficient control deviation, thus the prediction problem is actual for generating system optimal efficiency ensuring. Prediction techniques are based on mathematical statistics and analysis of acquired data. One of the methods is regression analysis, which describes relationship between outcome (dependent) value and one or more features (independent values or predictors). In other words, regression analysis
defines a function that approximates given values and generalize the certain case, so it is possible to find the outcome value for new predictor values.

2. Regression analysis

For regression analysis is purposed to solve problems of outcome prediction and feature influence evaluation on the dependent value, then it is reasonable to study its application to address the problem of TEI estimation [4]. Parameter influence can be defined numerically, so the higher its value the more important the predictor, therefore this variable should be considered when monitoring generating unit operation to ensure its higher efficiency.

In regression analysis the function that describes the relationship between outcome value and predictors can be presented as:

- linear;
- quadratic;
- cubic;
- logarithmic;
- exponential;
- logistic etc.

Commonly the regression function formula is defined by empirical and statistic data in such way that it fits the data with minimal error, which is mean square distance for each instance.

For number of parameters and signals in DCS informational field is great and not all of them are same influential on the outcome value, then it is reasonable to select ones that are most important to be considered when estimating TEI. The reason for this is that when monitoring thermal power plant generating unit better efficiency can be ensured with optimal regulation and control mechanism adjustment, which requires a heuristic approach, i.e. operator decision. Introduction of generating unit efficiency prediction and parameter selection allow making the feed-back channel to avoid necessity in manual adjustment.

When carrying out regression analysis the regression function, which general form is defined by (1), coefficients are found by least square method (LSM) as shown in formula (2).

\[
f(x) = b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_n \cdot x_n + a
\]

\[
\sum_i e_i^2 = \sum_i (y_i - f(x_i))^2 \rightarrow \min
\]

The disadvantages of classic regression analysis approach are necessity to choose the regression function that fits the data mostly heuristically and the assumption that data is linear separable, that limit application of regression analysis for operating calculations. To solve that problem application of machine learning methods is proposed. In paper [5] it is noted that application of artificial neural networks (ANN) allows solving multivariate regression analysis problem where data is not linear separable. The general ANN structure (figure 1) includes input values (features) X; weights W and bias b, which are dynamically updated when training to fit data; summing Z and activation A functions. The outcome value is output Y.
The distinctive feature of the technique is non-direct regression function determination and its coefficients, but step-by-step model weights adjustment during a certain number iterations - such a procedure is called training. To determine how weights should be updated at the of each iteration the error has to be calculated - cost function $L$ (3) and its partial derivative by weights $\frac{dL}{dw}$ (4) to be found. This training method is called gradient descent (GD). The purpose of training process in minimizing cost function of model to fit data most accurate - in case of regression analysis cost function is difference between predicted outcome $\hat{Y}$ and ground-truth one $Y$ given in dataset for corresponding value of independent variables $X$ [6].

$$L = \frac{1}{2} \sum_{i=0}^{m} (y_i - \hat{y}_i)^2$$  \hspace{1cm} (3),

where $m$ - dataset samples number.

$$w_i = w_i - \alpha \cdot \frac{dL}{dw_i}$$ \hspace{1cm} (4),

where $\alpha$ - learning rate.

When training and testing model performance how it generalizes data given dataset usually is split in training and testing datasets with proportion 80/20 respectively. Should be mentioned that solving problem of non-linear separable data requires at least one hidden layer in ANN model - a one layer ANN represents linear regression. For regression analysis with ANN the common activation functions are linear (5) and ReLU (6).

$$a(x) = x$$ \hspace{1cm} (5)

$$a(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$ \hspace{1cm} (6)

3. Parameter selection

Application of ANN regression model to predict thermal power plant generating unit performance which is relationship between gross efficiency and control parameters is proposed. For the number of hidden layers and their units are chosen heuristically, then it is proposed to build multiple models with various structure and evaluate their performance to select the optimal one. To train the model the historical operating parameters and TEI evaluation outcomes can be used. Once the model is fitted it

can be used to make predictions on new given data which was not presented in the training dataset. To rank the features by its influence on the outcome application of the permutation importance technique is suggested. It consists in shuffling feature values one-at-time and evaluating changes in outcome, how cost function suffered from the procedure. Once all feature importance are evaluated information about which parameters should be considered, be monitored thoroughly in order to ensure higher generating unit efficiency is given to operator, thus it is possible to automate pinpointing parameters which regulation should be optimized.

Conclusions
Application of the suggested machine learning techniques in computational stations allows automating most influential input signal selecting and predicting outcome values by solving regression problem with large number of features. The proposed approach provides heat and power object operation efficiency prediction and finding the most important parameters, what allows making the feed-back channel for parameter adjustment recommendations to operator, by doing so it provides elements of consulting system.

References
[1] Arakeljan E K and Galustjan M K 2012 Metodicheskie podkhody k raschotu godovykh tehniko-ekonomicheskikh pokazateley energeticheskogo oborudovaniya avtonomnogo kompleksa proizvodstva elektroenergii i tepla Novoe v Rossiyskoy energetike 5 16-28
[2] Arakeljan E K, Vasil’ev E D and Khurshudjan S R 2014 Aumomated control system issue and a possible way of its decision Vestnik MPEI 1 15-20
[3] Sultanov M M, Boldyrev I A and Gorban Y A 2019 Electrical generator unit technical and economic indexes parameter study 1st IEEE international youth conference on radio electronics, electrical and power engineering, REEPE 2019 (Moscow: Moscow Power Engineering Institute) p 870
[4] Draper N R and Smith H 1998 Applied Regression Analysis, Third Edition (New York: John Wiley & Sons, Inc.)
[5] Zivkovic Z, Mihajlovic I and Nikolic D 2008 Artificial neural network method applied on the nonlinear multivariate problems Serbian journal of management 4 143-155
[6] 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)
[7] Altman A, Tolosi L, Sander O and Lengauer T 2010 Permutation importance: a corrected feature importance measure Bioinformatics 26 1340-1347

Acknowledgements
The research is funded by Russian Federation public contract № 0720-2020-0025 "Technique development and method analysis for ensuring power system object security and competitiveness based on the digital technologies"

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