Machine Learning Approach Application for High-voltage Instrument Transformers Technical State Assessment

Alexandra Khalyasmaa[1,*], *Alina* Stepanova[1], *Daria* Shatunova[1] and *Vadim* Manusov[2]

[1]Ural Federal University, 620002 Mira Street 19, Russian Federation
[2]Novosibirsk State Technical University, 630073 Prospekt K. Marksa, Novosibirsk, Russian Federation

Abstract. This paper describes the possibilities of machine learning application in the tasks of technical state assessment of high-voltage instrument transformers. An analytical review of modern systems for technical state assessment of high-voltage equipment is presented, their advantages and disadvantages are described. A mathematical model of an automated system for assessing the high-voltage instrument current and voltage transformers based on gradient boosting over decision trees has been developed. The efficiency of the developed solution is proved using the example of analysis of a real distribution zone, which allows identifying the state of instrument transformers with an accuracy of 84%.

1 INTRODUCTION

The most important requirements for any high-voltage equipment at power plants and substations are reliability and operation safety, however current and voltage instrument transformers, which serve to monitor the main parameters of the electrical network, cannot always meet them.

According to the annual reports of electric grid companies in Russia, on average, about 30% of current and voltage transformers demonstrate degraded state and more than 30-40 years lifespan [1], therefore they are potentially dangerous (unreliable) nodes of the electrical network. The main danger in this case is linked with the consequences that can be caused by serious faults of obsolete types of oil current and voltage transformers. For instance, it could be breakdown of the insulation with an explosion resulting in:

− hitting and disabling neighboring equipment within 200 meter radius;
− oil release into the atmosphere which leads to a fire with a sufficiently wide burning area.

It is worth noting that the damageability of current and voltage instrument transformers still remains quite high.

To solve the tasks of improving the reliability of electrical networks, current and voltage transformers are subject to mandatory state diagnostics which are regulated by reference documentation. However, today not all types of technical diagnostics of the state of current and voltage transformers in Russia have specialized separate regulatory documents or criteria for their parameters interpretation. Most often, their diagnostics is carried out in accordance with the requirements of regulatory documents for power transformers which in the opinion of the authors of this paper is incorrect due to fundamentally different purposes of equipment and their different operating conditions and modes.

In addition, the accuracy of current and voltage transformers also determines the selectivity of the operation of power system protection influencing the reliability of the power system.

2 Technical state assessment today

Currently, on-line and off-line diagnostics methods are used to assess the technical state of high-voltage equipment at power stations and substations.

On-line diagnostics methods are meant to perform monitoring the state of the equipment under voltage in order to obtain information on its current state allowing its subsequent processing. It is generally accepted because of the complexity of implementing this type of diagnostics and the need for processing a large amount of data that the monitoring system is installed on one piece of equipment which is usually the most expensive, for example, on power transformers as implemented by companies such as ABB [2], Siemens [3], LumaSense [4].

It is worth noting that the existing foreign systems for assessing the technical state of instrument transformers, for example developed in the USA and Europe, also cannot be applicable in Russia as there are large differences in the regulatory documentation for equipment maintenance, diagnostics, configuration and operation. Their substantial processing is required to be implemented in power companies of Russia.

Off-line diagnosis implies an analytical analysis of the data obtained during the tests and methods of non-destructive monitoring of the equipment state. These include, for example, the expert system for assessing the

---

* Corresponding author: lkhalyasmaa@mail.ru

technical state "Diagnostics +"[5], based on the Bayesian method, or the expert system for diagnosing oil-filled equipment "EDIS Albatross" [6], operating on the basis of neural networks.

The feature of existing systems of both on-line and off-line diagnostics of the state of high-voltage equipment is mainly their focus on solving a specific problem of a certain owner of power grid assets (for specific network topology and equipment). In this case, the systems are almost completely inflexible and cannot be used in their original form at other sites without processing. Also, such systems use different-scale information of different accuracy without taking into account the dynamics of the criteria for assessing the technical state of the equipment, according to the latest time data snapshot, which can lead to errors in identifying the state of equipment.

Therewith, there is a number of disadvantages not only in the diagnostics systems, but also in the approaches to assessing the state of the equipment, namely:

- it has been scientifically proven that oil analysis in the absence of control under voltage remains practically the only way to detect developing damage, but at the moment Russia has not developed standards for instrument transformers on the content of gases dissolved in oil;
- heat monitoring is more widespread, but is often performed without taking into account the design features of instrument transformers and is not always thorough;
- as a rule, only tests and measurements of characteristics prescribed by the standards are carried out at substations, without analyzing the results, i.e. without specifically diagnosing the state of equipment insulation.

Furthermore, today the world power industry including the Russian Federation is on the path of digitization which leads to creation of new types of equipment such as fiber-optic electronic current transformers and electronic voltage transformers based on capacitive or non-inductive resistive high-voltage divider.

Using optical methods for measuring current allows obtaining measured values forthwith in digital form, and the applied voltage-meter circuit makes it possible to significantly improve the measurement accuracy and reduce errors. At the moment approaches to assessing the state or this type of equipment have not yet been developed, but they are already being actively used at various energy facilities.

All of the above arguments underline the need for developing a completely new approach to assessing the current technical state of current and voltage instrument transformers which would allow increasing the accuracy of identification of developing faults under uncertainty. The authors of this paper have developed a mathematical model of an automated system for assessing the technical state of high-voltage instrument current and voltage transformers based on machine learning methods that

can assess the state of not only electromagnetic, but also optical ones.

3 Mathematical methods for assessing the technical state

3.1 Artificial intelligence methods in state assessment tasks

In this paper, the authors present a developed mathematical model of an automated system for assessing the technical state of high-voltage instrument current and voltage transformers based on machine learning methods.

Existing systems for assessing the state of the equipment mainly analyze the state of power transformers, circuit breakers and disconnectors making diagnosing the state of current and voltage transformers impossible.

In addition, existing systems have an identification accuracy of 70-75% while the developed system has an accuracy of 80-85%, i. e. the percentage of erroneous diagnostics is not 25-30%, but 15-20%, which increases the reliability of the power supply system.

The use of standard mathematical methods on a rigid algorithmic basis for solving the problem of assessing the technical state of power grid facilities and making decisions on their further operation may give unsatisfactory results. The reasons are the complexity of the objects being analyzed and the fact that solving the posed problems requires operating with a large amount of input data of both numerical and linguistic format, which are characterized by uncertainty, incompleteness and the lack of formal structuring.

Solving the set tasks requires methods that can perform such a function of human intelligence as choosing the optimal solution based on the experience gained earlier and rational analysis of all available information on the object of study. From the point of view of the mathematical apparatus based on which the existing intelligent systems in the power industry function, the following ones can be called the most common [7]:

- – artificial neural networks (ANNs) – a mathematical model of biological neural networks and human brain neurons describing the principles of their organization and functioning [8];
- – heuristic search systems (genetic algorithms) – algorithms used to solve optimization and modeling problems based on genetic processes similar to biological organisms [9];
- – knowledge-based systems (expert systems; inference systems), i.e. systems built on rules stored in a knowledge base, which are used to find solutions and conclusions from the facts based on a set of initial facts [10].

Despite the fact that there are many methods of intelligent analysis, they are practically not used in the actual operating systems for assessing the technical state. The fact is that most of intelligent analysis methods

require the development of a mathematical model of equipment, which for each piece of equipment is a separate task under uncertainty sometimes being difficult to implement due to the many processes of different nature that occur in the equipment itself [11].

3.2 Machine learning in state assessment tasks

In the developed system within the framework of the presented research it was proposed to use the gradient boosting over decision trees as a mathematical algorithm, which is the most effective way to optimize the objective function for the problem being solved. During the model formation the basic algorithms are summed, with each subsequent algorithm eliminating the errors of the previous ones by finding the maximum decrease of the objective function. Thus, the gradient boosting sequentially builds the model during the gradient descent in the algorithm space.

In the problem being solved, the algorithm is used to assess the state of instrument transformers when analyzing various test and diagnostic results.

The idea of the boosting approach is to combine weak (with low generalizing ability) functions that are built during an iterative process, where at each step a new model is trained using the error data of previous ones. The resulting function is a linear combination of basic, weak models.

Let training a composition of N decision trees be necessary:

$$a_N(x) = \sum_{n=1}^{N} b_i(x), \qquad (1)$$

where $a_N(x)$ is the composition of decision trees, $b_n(x)$ is the decision tree included in the composition.

For the classification task a loss function is formed which looks as follows [1]:

$$L(y,z) = y\ln\big(1+\exp(-z)\big) + (1-y)\ln\big(1+\exp(z)\big), \qquad (2)$$

where $L(y,z)$ is the loss function, y is the true answer, z is the algorithm forecast.

The composing begins with initialization, i.e. with building a basic algorithm $b_0(x)$. A basic algorithm can be $b_0(x) = 0$, which always returns zero.

Training basic algorithms occurs sequentially. Let by some time of training N-1 algorithms $b_0(x),...,b_{N-1}(x)$ their composition be as follows:

$$a_{N-1}(x) = \sum_{n=1}^{N-1} b_n(x), \qquad (3)$$

After that, one more algorithm is added to the current composition $b_N(x)$. This algorithm is trained to minimize the composition error on the training sample [**1**]:

$$\sum_{i=1}^{l} L(y_i, a_{N-1}(x_i) + b(x_i)) \to \min, \qquad (4)$$

The first step is to solve a simpler task: determine which values of $s_1...s_l$ the algorithm $b_N(x_i) = s_i$ should take on training sample objects to minimize error:

$$\sum_{i=1}^{l} L(y_i, a_{N-1}(x_i) + s_i) \to \min, \qquad (5)$$

where $s_1...s_l$ is the shift vector.

In other words, it is necessary to find a shift vector S that minimizes the function $F(s)$. Since the direction of the fastest decrease of the function is given by the direction of the ant gradient, it can be taken as a vector S:

$$s = -\Delta F = \begin{pmatrix} -L(y_1, a_{N-1}(x_1)) \\ .... \\ -L(y_l, a_{N-1}(x_l)) \end{pmatrix} \qquad (6)$$

The components of the shift vector s, in fact, are the values that the new algorithm $b_N(x)$ must accept on the objects of the training sample in order to minimize the error of the constructed composition.

4 Training sampling and calculation example

Within testing the developed system, the technical state of current transformers was assessed. A database created based on the results of diagnostics of oil-filled power equipment, operational experience, and acceptable parameters, included a sample of technical indicators for the last 20 years for 13 single-type current transformers.

The first priority for machine learning methods is to convert the data into a table view and create a database. This process is the most difficult and time-consuming.

Tabular representation is the most common representation of data in the field of machine learning and data mining. The rows of the table are separate objects (observations). The columns of the table contain independent variables (attributes), denoted by X, and dependent (target) variables, denoted by y.

The metric is needed to evaluate the results of machine learning algorithms. In this case, the average percentage error Q [%] of the algorithm a on the sample X^l was used as a metric, which is calculated by the formula:

$$Q(a, X^l) = \frac{1}{l} \sum_{i=1}^{l} |a(x) \neq y^*(x)| \cdot 100\%, \qquad (4)$$

where $|a(x) \neq y^*(x)|$ is the error indicator; $y^*(x)$ is the target dependence; l is the number of observations.

Under testing the developed system, a training sample based on the results of oil-filled power equipment diagnosing, operating experience, and permissible parameters limits was formed including a sample of the following 11 technical indicators (parameters) since 1970 for 13 single-type 110 kV current transformers (TFND type):
- transformer oil specification;
- transformer refill oil specification;
- the value of the dielectric loss tangent at 20°C;
- the value of the dielectric loss tangent at 70°C;
- the value of the dielectric loss tangent at 90°C;
- aqueous extract reaction;
- water soluble acids;

– mechanical impurities;
– closed flash point;
– acid number;
– water content.

Each set of parameters was corresponded by its state of the equipment identified by the specialists in the laboratory of oil-filled equipment diagnostics and had the following gradation:

– operational state D_1 when the transformer fully meets all technical requirements (serviceable);
– a faulty but operational state D_2 when only those properties of a transformer that characterize its ability to perform specified functions meet technical requirements;
– limit state D_3 in which its further operation is unacceptable or unfeasible for various reasons.

In this study to assess the technical state of current transformers, they were determined to belong to one of the following three states using gradient boosting over the decision trees and the average error of state identification in percent [%] analysis in comparison with the states determined by the laboratory of oil-filled equipment.

The size of the training sample, developed test sample included 446 pairs of values.

The accuracy of the technical state assessment also depends on the depth of the decision trees and their number in the composition of the algorithms. The depth of the tree is determined by the greatest depth of the leaf node.

Calculations revealed that the optimal accuracy of the technical state assessment algorithm is achieved when a depth of decision trees is 5 and their number in the composition is 6.

The main results of state identification based on gradient boosting over decision trees are presented in Table 1.

Table 1. The main characteristics of the developed model

Characteristic	Values
The number of pairs in the training sample, pcs.	446
Optimal depth of decision trees, pcs.	5
Optimal number of decision trees in the composition, pcs.	6
Average training error, %	0.88
Average testing error, %	0.84
Type I errors, pcs.	44
Type II errors, pcs.	28

In addition, the study analyzed not only the size of errors but also their type and identified among them the errors of the first ("false defect") and the second ("defect skip") type.

These errors are due to the fact that the input value of the parameter has dispersion on the boundary regions and the same values can correspond to one and the other state. The type I error leads to additional preventive work, and the type II error entails costs including not only emergency repairs but also equipment depreciation, etc.

The analysis performed describes the system as fairly reliable, and in this case, 16% of errors in determining the state of current transformers can be considered a good result.

5 Conclusions

Within this study a mathematical model of an automated system for assessing the technical state of high-voltage current transformers based on machine learning methods was developed and tested.

It was also proved that the task of assessing the technical state of current and voltage instrument transformers is reduced to the standard machine learning task, multiparameter classification, or pattern recognition by precedents based on a pre-formed training sample.

To solve the task of pattern recognition of the power circuit breakers technical state the XGBoost machine learning algorithm was used, which is based on the decision tree algorithm.

The possibility and efficiency of using machine learning (in comparison with existing approaches) in assessing the technical state of equipment under uncertainty using the example of analyzing the state of current transformers in a large power center of the Sverdlovsk region is proved.

The ways to solve the problem of improving the system for assessing the technical state of power grid equipment based on diagnostic information in accordance with current requirements and taking into account formalized knowledge and experience of experts are determined.

References

1. PJSC "FGC UES". Annual Report for 2016. URL: http://www.fsk-ees.ru/shareholders_and_investors/ disclosure_of_information/annual_reports/ (Date of circulation: 13.10.2018).
2. J. Vines, B. Banh, C. Stiegemeier, P. Patel, L. V. Cheim. Transformer health in the real world. URL: https://library.e.abb.com (Date of circulation: 13.10.2018).
3. Energy Management Division. Article No. EMSG-B10006-00-7600. URL:http://m.energy.siemens. com/MX/pool/hq/services/power-transmission-distribution/condition-monitoring/ (Date of circulation: 18.10.2018).
4. ThermalSpection 724 system. URL: https:// www.lumasenseinc.com/EN/products/temperature-measurement/thermal-imagers-systems/thermal-imaging-systems/thermalspection-724-system/ther malspection-724.html. (Date of circulation: 24.09.2018).

5. System for estimation of electrotechnical equipment "diagnostics. URL: http://transform.ru/ diagnostika.shtml (Date of circulation: 24.09.2018).

6. I.V. Davidenko, T. Kuzina, Analysis of the modern methods of the power transformers health index calculation. Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering Conference, ElConRus 2017. Institute of Electrical and Electronics Engineers Inc., p. 1491-1495 5 p. 7910856.

7. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems. Publisher: O'Reilly Media. 2017. 572 p.

8. S.J. Russell, P. Norvig. Artificial Intelligence: A Modern Approach. Publisher: Pearson; 3 edition. Authors:. 2009. 1152 p.

9. M. Mitchell. An introduction to genetic algorithms. Publisher: MIT Press. 1996. 221 p

10. G. James, D. Witten, , T. Hastie, R. Tibshirani. An Introduction to Statistical Learning. Publisher: Springer-Verlag. 2013. 572 p.

11. Khalyasmaa, A. Integral assessment of power network equipment operational risks: Special aspects. Lecture Notes in Electrical Engineering. 2019. P. 280-286