The Research of Anomaly Detection Method for Transformer Oil Temperature Based on Hybrid Model of Non-Supervised Learning and Decision Forests

Fei Xiao¹, Xiwu Leng², Kang Ye¹, Youlin Hu¹, Xiongli Li³, Licheng Zhu³

¹ State Grid Shanghai Electric Power Company Power Dispatch Control Centre, Shanghai, P.R. China
² National Power Dispatch Communication Centre, Beijing, P.R. China
³ Tellhow Software Co. Ltd., Nanchang, P.R. China
E-mail: xiaofei@sh.sgcc.com.cn; xiwu-leng@sgcc.com.cn; yekang_yk@sh.sgcc.com.cn; 3311481@qq.com; 66653752@qq.com; zhulc@tellhow.com

Abstract. The anomaly detection of transformer’s oil temperature is critical and valuable issue for the safe operation of transformers and power system. In terms of the defects of traditional anomaly detection approaches of transformer’s oil temperature, such as high investment, poor generality, and non-real time, this paper proposed a hybrid model with non-supervised learning and decision forests method to detect anomaly of transformer’s oil temperature. Based on non-supervised clustering algorithm, firstly, the clusters of transformers’ working conditions are explored from big data sets of transformers. After that, the abnormal temperature threshold value of each cluster is deduced by hypothesis tests method and utilizes to tag anomaly in data sets of working conditions. Finally, the data sets with anomaly tags are inputted into random decision forests model to construct the classifier of abnormal oil temperature and generate the rules for anomaly detection. This method was validated by empirical data of main transformer in Shanghai, and the results represented its conspicuous competitive advantages to traditional oil temperature anomaly detection methods in the factors of real-time and accuracy.

1. Introduction
Transformer [1]-[3] is the core component of energy transmission in power system, which is expensive and technically complex, and the oil temperature of transformer is closely related to the service life of transformer [4]-[6]. Excessively high oil temperature will seriously shorten the service life of transformer. The main reasons for excessively high oil temperature involve the insulation damage of the metal parts of the transformer, winding inter-turn short-circuiting, overload of transformer, failure of transformer cooling system and so on. At present, the classical method for the calculation, prediction and abnormal analysis of transformer oil temperature are generally based on transformer modified heat path model technology, transformer oil chromatographic discrimination technology and top layer oil temperature anomaly detection technology. The heat path model calculation technique is a traditional method for anomaly detection of transformer oil temperature. It is mainly based on the heat path model calculation of the main transformer equipment [7]-[9], which establish a heat path for the heat transfer process of the transformer and judge the abnormal oil temperature. Oil chromatographic anomaly detection technique, which analyse the content of measurable gas in transformer oil by chromatographic analyser to detect anomaly of oil
temperature, is widely used at present [12], [13]. The anomaly detection technology of oil temperature at the top layer is to identify the abnormal oil temperature based on the sampling data of the top oil temperature of transformer and the characteristic model of oil temperature. A semi-physical model for short-term prediction about the peak oil temperature of the top layer was constructed, and the oil temperature was forecasted with consideration on the influence of the operation of the cooler and the abnormal working condition of the transformer [10], [11]. Despite that these classic anomaly detection methods of oil temperature already have many outstanding achievements, which still has the problems involving high costs, difficulty of configuration, and non-real time, how to be more efficient in anomaly detection of transformer oil temperature is still an important research topic. In the era of big data, this paper proposes a non-supervised hybrid model method for oil temperature anomaly detection, which bases on clustering, hypothesis testing and decision forest algorithm [14], [15]. Compared with the classical temperature anomaly detection methods, this method has the advanced features of low investment, pervasive generalization, and real time.

2. Anomaly detection hybrid model of oil temperature based on non-supervised learning and decision forests

According to the practical experiences of oil temperature anomaly detection, which summarized by experts of State Grid Shanghai Electric Power Company, the expert rules was manifested as following: 1) When a transformer is in same working condition, the abnormal temperature is small probability event. 2) When a transformer is in the same temperature and with different working conditions, the temperature may be normal or abnormal.

Moreover, the relevant data set of anomaly detection of oil temperature can be considered as a five-element information system, which composed of the objects, the attributes of relevant working conditions, the anomaly tagging attributes of oil temperature, the domain of attributes’ value and the relationship of value function. The oil temperature anomaly tagging information system can be formally described as $\text{IS} = \{U, C, D, V, f\}$. In this system, $U = \{S_1, S_2, L, S_m\}$ means the objects collection for working condition records of $m$ transformers. $C = \{C_1, C_2, L, C_n\}$ means the attributes collection for relevant $n$ working condition records, in which oil temperature is also an attribute. $D$ represents a tagging attribute of oil temperature anomaly, $V$ represents the value domain of relevant working conditions attributes and oil temperature anomaly tagging attribute. $f = U \times (C / D) \rightarrow V$ represents the functional relationship between the attributes of different objects and their values, such as the attribute values of object $S_i$ can be recorded as $C_{i1}, C_{i2}, L, C_{in}$ and $D_i$.

In terms of null original values of $D$ attribute of all objects in oil temperature anomaly tagging information system $\text{IS}$, that is $D_i = \emptyset$, $i \in m$, the main purpose of this paper is to determine how to obtain the value of the oil temperature anomaly tagging attribute $D$, according to the $n$ working condition attribute sets $C = \{C_1, C_2, L, C_n\}$ of the transformer at a certain time. Based on the expert rules and the oil temperature anomaly tagging information system, the anomaly detection method in this paper, firstly, was to cluster working conditional data set of transformers via non-supervised learning method. Secondly, abnormal oil temperature in each condition clustering with small probability is tagged, and the granular threshold value of anomaly detection was determined by hypothesis testing strategy. Finally, abnormal oil temperatures were tagged by decision forest algorithm model from big data set of relevant transformers’ working conditions.

2.1. Machine learning strategy for anomaly detection hybrid model of oil temperature

In this paper, based on oil temperature anomaly tagging information system $\text{IS} = \{U, C, D, V, f\}$ and experience of industrial experts, anomaly detection of oil temperature is taken consideration as typical non-supervised learning problem, and a novel anomaly detection hybrid model with clustering,
hypothesis testing and decision forest algorithm is proposed. Figure 1 as following describes the work flow and steps of this hybrid model.

![Flow chart of anomaly detection hybrid model of oil temperature](image)

**Figure 1.** Flow chart of anomaly detection hybrid model of oil temperature

The stages of this hybrid model are as follows:

Stage 1: Data preprocess. Historical data of relevant working conditions, which include winding temperature, oil temperature, environmental temperature, active power, reactive power, current etc, were preprocessed. The preprocessing works included data cleansing, data wrangling and data discretization.

Stage 2: Feature Extraction. big data set of relevant working conditions were clustered via K-means algorithm [16], [17]. Different types of working conditions were explored, and the probability density distribution of oil temperature in each cluster of working condition were computed. Furthermore, the small probability samples of each cluster are tagged as anomaly of oil temperature by granular threshold value via hypothesis testing, and a two-dimensional decision information table for anomaly detection can be constructed.

Stage 3: Classifier Construction. The classifier of oil temperature was constructed to detect the abnormal oil temperature via decision forests model [18]-[20]. The decision forests model was trained firstly by the data in two-dimensional decision information table and utilized to detect anomaly of oil temperature after training stage from real-time data of transformers’ working condition.

Stage 4: Classifier Evaluation. The classifier of oil temperature was evaluated, and the granular threshold value was updated if precision and recall test of model did not reach the preset. The gradient pattern adjustment of threshold value was repeated to execute from stage 2 to stage 3 until the preset was satisfied.

2.2. The methodology of anomaly detection hybrid model of oil temperature
The anomaly detection hybrid model of oil temperature involved clustering of transformers’ working conditions, anomaly tagging of oil temperature, decision forests, and hypothesis testing. However, based on K-means clustering about transformers’ working conditions, the final classifier was deduced from decision forests algorithm, which was evaluated and optimized via hypothesis testing. The critical algorithms and methods of model was illustrated in algorithm 1 as following:

**Algorithm 1:** anomaly detection hybrid model for oil temperature based on K-Means clustering and decision forests algorithm.

**Algorithm input:** historical data of relevant transformers’ working condition, $k$ values of K-means clustering, granularity threshold value $\gamma$ of anomaly.

**Algorithm output:** efficient classifier of abnormal oil temperature.

**Step 1:** according to historical data of relevant transformers’ working condition, K-Means clustering algorithm is applied to obtain $k$ types of data sets, and $j$ cluster are described as $U_j$, $j \in [1, 2, \ldots, k]$.

**Step 2:** based on the clusters from Step 1, the conditional probability of oil temperature in each cluster is computed, and the conditional probability of $j$ cluster is described as $|U_j|/|U|$.

**Step 3:** conditional probability value of every oil temperature in each cluster was compared with the granularity threshold value $\gamma$. If $|U_j|/|U| \geq \gamma$, the oil temperature samples are tagged as normal in its cluster. Otherwise, when $|U_j|/|U| < \gamma$, the oil temperature samples are tagged as anomaly in its cluster. Finally, two-dimensional decision information system $IS = \{U, C, D, V, f\}$ was established via anomaly tagging, which based on granular threshold value.

**Step 4:** the two-dimensional decision information system $IS$ in Step 3 was input into decision forests as training data set to obtain a strong classifier of transformers’ oil temperature. Thus, when online data of transformers’ working conditions was input into this strong classifier, the normal or abnormal oil temperature can be tagged from the data set.

**Step 5:** the precision and recall rate of the strong classifier in Step 4 was evaluated to test its efficiency. If the precision and recall rate was fails to reach the preset value, the granular threshold value $\gamma$ of anomaly was updated to $\gamma = \gamma + \alpha$ and $\alpha$ was adjusted as the greedy algorithm. The optimized process of the strong classifier was iterated from Step 1 to Step 4, until the preset value was satisfied.

**3. Experiments and analysis**

In the experiment, the data of no. 4 transformer of substation in Shanghai from December 2014 to January 2017 were utilized to validate the efficiency of anomaly detection hybrid model of transformers’ oil temperature. There were 18 rational attributes of transformers’ oil temperature to be input into model, which include temperature of environment, temperature of A winding, active power from high voltage side, active power from middle voltage side, active power from low voltage side, A phase current amplitude from high voltage side etc. The data from year 2015 to year 2016 were leveraged to train the anomaly detection hybrid model, and the final classifier of model was applied to explore abnormal oil temperature from data set of year 2017. A synthetical index with precision and recall rate was used to evaluate the efficiency of final classifier of anomaly.

Figures 2 and 3 demonstrate the experimental results of the anomaly detection hybrid model, which tag the abnormal oil temperature in March 2017 and July 2017 with red spot mark. Obviously, the abnormal oil temperature of the main transformer equipment had been tagged accurately in figure 2 and 3. In figure 2 and 3, the above subgraph represents the values of 3 main attributes in a month, including the temperature of environment (red line), the current amplitude of phase A from high voltage side (blue line), and the temperature of Phase A winding (green line). In figure 2 and 3, the below subgraph represents the value of oil temperature with anomaly tagging (red spot).

The experimental results were compared with the records in EMS and analyzed by the experts. The precision of oil temperature anomaly tagging is 90%, and the recall rate of oil temperature anomaly tagging is 89%. However, the precision of classical transformers’ oil chromatographic anomaly
detection method has only 86%. The method proposed in this paper is better than the classical oil chromatographic anomaly detection method in the factor of precision and efficiency.

![Figure 2](image1.png)

**Figure 2.** Anomaly tagging of oil temperature in March 2017

![Figure 3](image2.png)

**Figure 3.** Anomaly tagging of oil temperature in July 2017

### 4. Conclusion
The anomaly detection of oil temperature can explore the abnormal state of working condition of transformer, and provide clues to support the prediction of transformers’ defects. According to the research of classical anomaly detection technology of oil temperature, a novel anomaly detection hybrid model of transformers’ oil temperature, which combined non-supervised learning strategy with decision forests algorithm, was established from the electrical big data of transformers’ working conditions in this paper. The critical approach was to tag anomaly of oil temperature with non-supervised learning methods from the aspect of statistical meanings, and to adjust the threshold value of tagging anomaly via hypothesis testing method, which optimized the anomaly detection method of oil temperature adaptively. Furthermore, the results of experiment, in which the method was implemented to detect abnormal oil temperature of some real transformers in Shanghai, have validated
the efficiency and accuracy beyond classical methods mentioned before. In addition, this novel anomaly detection methods of oil temperature, did not need extra investment for specific equipment, and can be suitable for heterogeneous types of the transformers oil temperature alarm system. Thus, it is worth to promote for its sound commercial value and social value.

5. Acknowledgments
I really appreciate the support and assistant of every expert and colleague from State Grid Shanghai Company for this project. Their advices and knowledge made this method discussed in the paper to be feasible and implement. However, I also thank for the help of Dr. Yu, Dr. Xu and Dr. Zhao from Nanchang University and Tellhow Company.

6. References
[1] Lu Z G, Zhao G, Yang D L, et al. 2016. Overview of Research on Power Electronic Transformer in Distribution Network. Proceedings of the CSU-EPWA. 28(5). p48-54.
[2] Anonymous. Future development trend of distribution transformer industry in China[J]. Transformer, 2012. 49(2). p51-51.
[3] Han L, Wang M Y, Zhou Y X, et al. 2015. Effect of Oil Gap Spacing on Characteristics of Partial Discharge of Oil-Paper. Transformer. 52(8). p24-27.
[4] Lai R H. 2017. Analysis of substation main transformer status maintenance technology status quo and development trend. Science and Information Technology. (2).
[5] Wei Y B, Wang D H, Han L F, et al. 2015. A novel method for discharging fault diagnosis and location of oil-immersed power transformers based on MIA. Power System Protection and Control. 43(21). p41-47.
[6] Li S G, Xue H, Li Z, et al. 2017. Fault diagnosis of mine-used transformer based on optimized fuzzy Petri net. Industry and Mine Automation. 43(5). p54-57.
[7] Wei B G, Huang H, et al. 2012. Algorithm for Transformer Top-Oil Temperature and Winding Hot-Point Temperature Based on Modified Thermal Circuit Model. East China Electric Power. 40(3). p0404-0407.
[8] Chen W G, Su X P, Chen X, et al. 2011. Influence Factor Analysis and Improvement of the Thermal Model for Prediction Transformer Top Oil temperature. High Voltage Engineering. 37(6). p1329-1335.
[9] Wang Y Q, Yue G L, He J, et al. 2014. Study on Prediction of Top Oil Temperature for Power Transformer Based on Kalman Filter Algorithm. High Voltage Apparatus. (8). p74-79.
[10] Chen J M, Wu Y, Zhu H B, et al. 2015. Research and Application of Transformer Top Oil Temperature Short-term Prediction Model. Electrotechnical Application. (22). p89-93.
[11] Du S Y, Wang H B, Li F. 2015. Transformer oil temperature abnormal state identification method. Electrotechnical Application. (82). p859-862.
[12] Zhou D J. 2015. Analysis of Abnormal Chromatogram in Large Transformer Oil and its Technology Principle for Investigation. Electric Power Science and Engineering. 31(1). p31-37
[13] Gao S G, Wang X L, Li Q M, et al. 2014. Outliers Detection and Distribution Characteristics of the Transformer DGA Data Based on MCD Robust Statistics. High Voltage Engineering. 40(11). p3477-3482.
[14] Zhuang C J, Zhang B, Hu J, et al. 2016. Anomaly Detection for Power Consumption Patterns Based on Unsupervised Learning. Proceeding of the CSEE. 36(2). p379-387
[15] Wang X, Liu X Q, Song S L, et al. 2014. Unsupervised Learning Algorithm for Abnormal Behavior Detection. Opt-Electronic Engineering. 41(3). p43-48
[16] Naveen A, Velmurugan T. 2015. Identification of calcification in MRI brain images by k-means algorithm. Medical Mycology.
[17] Bhowmik T K, Parui S K, Kar M, et al. 2016. Segmental K-Means Algorithm Based Hidden Markov Model for Shape Recognition and its Applications. International Journal of Human-Computer Studies. 88(38). p38-50.
[18] Afanador N L, Smolinska A, Tran T N, et al. 2016. Unsupervised random forest: a tutorial with case studies. Journal of Chemometrics. 30(5).
[19] Wu X Y, He J H, Zhang P, et al. 2015. Power System Short-term Load Forecasting Based on Improved Random Forest with Grey Relation Projection. *Automation of Electric Power Systems*. 39(12), p50-55.

[20] Zhao T, Wang L T, Zhang Y, et al. 2016. Relation Factor Identification of Electricity Consumption Behavior of Users and Electricity Demand Forecasting Based on Mutual Information and Random Forests. *Proceeding of the CSEE*. 36(3), p604-614