Advances in Computer, Signals and Systems (2021) 5: 42-47
Clausius Scientific Press, Canada
DOI: 10.23977/acss.2021.050106
ISSN 2371-8838

Transformer Fault Diagnosis Based on Stacking-Ensemble Meta-Algorithms

Yan Wang[1,a,*], Liguo Zhang[2,b]

[1]Dept. of Computer, North China Electric Power University, Hebei Baoding, China
[2] College of Information Science & Technology, Agricultural University of Hebei, Baoding, China
[a]Wangyan1206@126.com, [b]zhangliguo2006@126.com
**Corresponding author*

Keywords: Dissolved gas analysis; transformer fault diagnosis; Stacking-Ensemble; classification algorithm

Abstract: Compared with the method of establishing a single classifier for diagnosis, ensemble learning can combine multiple classifiers to achieve stronger generalization ability. This paper proposed a transformer fault diagnosis method based on Stacking Ensemble multiple classifiers, which can detect the transformer's internal fault by using its DGA data. The proposed model is consisted of two sections. The first section includes five diagnosis models: Random Forest Classifier, AdaBoost Classifier, Gradient Boosting Classifier, SVM and Extra Trees Classifier. The second section use XGB Classifier as final Meta-Classifier model to classify the faults of transformers by using all the base level model diagnosis results as input. The diagnosis accuracy of the proposed method is 83.3%, which is better than other single Classification method.

1. Introduction

Transformers are vital components of power systems as they serves as the connection of transmission and distribution networks at different voltage levels and their failure disrupts the use of electrical energy [1-4]. Several types of faults could jeopardize the reliable operation and continuous power supply of the apparatus. Therefore, early diagnosing an incipient fault is essential and effective in avoiding hazardous operating conditions and minimizes downtime cost.

DGA is an effective method for transformer fault diagnosis. It mainly detects the dissolved gas composition in transformer oil, including H2, C2H2, C2H6, C2H4, CH4. Many DGA data-driven fault diagnosis methods have been widely used [5,6]. J. Dai presents a new transformer fault diagnosis method based on deep belief networks (DBN). However, the intelligent diagnosis model based on single algorithm, such as BP, Random Forest Classifier, RVM, GP algorithm, usually can't get high accuracy.

Ensemble learning helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model [2,4]. The idea of stacking is to learn several different weak learners and combine them by training a meta-model to output predictions based on the multiple predictions returned by these weak models. In this paper, the two level transformer stacking diagnosis models are built. Which include base models and a Mental-classifier model. There are four diagnosis models used as base level models, namely Random Forest Classifier, AdaBoost Classifier, Gradient Boosting Classifier, SVM and Extra Trees Classifier. In the second section, XGB Classifier is used as final Meta-Classifier model to classify the faults of transformers by using all the base level model diagnosis results as input.

The rest of the paper is organized as follows: introduce base algorithms: Random Forest Classifier, AdaBoost Algorithm, Boosting algorithm. Then, discuss how to build and train the diagnosis model. In the third section, present the diagnosis results and give a real application. Finally in the last section there will be the conclusion.

2. Basic Classifiers

2.1 Random Forest Classifier algorithm

Random Forest often also called random decision forests, which is an ensemble of tree-structured classifiers (Fig.1). Usually it is often a collection of hundreds to thousands of trees, where each tree is grown using a bootstrap sample of the original data.Every tree of the forest gives a unit vote, assigning each input to the most probable class label [8].

The construction of Random Forest is described in the following main steps:

1) Draw k tree bootstrap samples from the original data set and the size of original data also is k.

2) Grow a tree for each bootstrap data set. Given the features number of each sample is M. At each node of the tree, Use information gain ratio method to select features select one of m(m<M) try features for splitting. Grow the tree so that each terminal node has no fewer than node size cases.

3) Aggregate information from k classification trees, and use voting to make new data classification predictions.

4) Compute an out-of-bag (OOB) error rate by using the data not in the bootstrap sample.

2.2 Boosting Algorithm and AdaBoost Algorithm

Boosting algorithms are a set of the low accurate classifiers to create a highly accurate classifier [9]. The basic idea of Boosting is to train the models serially instead of in parallel. AdaBoost Algorithm is the best representative lifting algorithm. By reducing the weight of the paired examples in each round and increasing the weight of the wrong examples, it makes the classifier gradually improved in the iterative process, and finally all the classifiers are linearly combined to obtain the final classifier. The AdaBoost algorithm flow is as follows:

(1) Initialize the training data weight distribution and each sample is given the same weights: $w_i = 1/N$. The weight distribution of the sample value is $D_1(i)$:

$$D_{1(i)} = (w_i, w_i, \dots w_i) = \left(\frac{1}{N}, \dots, \frac{1}{N}\right) \tag{1}$$

(2) Iterate:$\mathrm{t} = 1, \dots, \mathrm{T}$

(a) Select the classifier h (with the lowest error rate currently) as the t-th classifier H_t and calculate $h_t: X \to \{-1,1\}$.The error of h_t on distribution D_t is as follows:

$$e_t = p(H_t(x_t) \neq y_i) = \sum_{i=1}^{N} w_{ti} I(H_t(x_i) \neq y_i) \tag{2}$$

(b) Calculate the weight of the classifier in the final classifier

$$\alpha_t = \frac{1}{2} ln\left(\frac{1-e_t}{e_t}\right) \tag{3}$$

(c) Update the weight distribution of the training sample value D_{t+1}:

$$D_{t+1} = \frac{D_t(i)\exp(-\alpha_t y_i H_t(x_i))}{Z_t} \tag{4}$$

Here, Z_t is the Normalization constant: $Z_t = 2\sqrt{e_t(1-e_t)}$

(3) At last, combine each classifier according to the weight of the weak classifier:

$$f(\mathrm{x}) = \sum_{t=1}^{T} \alpha_t H_t(x) \tag{5}$$

Then, a highly accurate classifier is got:

$$H_{final} = sign\big(f(x)\big) = sign\left(\sum_{t=1}^{T} \alpha_t H_t(x)\right) \tag{6}$$

Here, (X_1,y_1), $(X_2,y_2)\ldots(X_n,y_n)$ are the train data set and $y_i \in \{1,-1\}$ denote the sample class. $D_t(i)$ is the weight distribution of the sample value, w_i is weight for *i*-th sample. *h* denotes weak classifier. *H* denotes basic classifier. H_{final} is final highly accurate classifier. *e* is error rate. α_t is the weight of weak classifier.

3. Stacking Diagnosis Model

3.1 Model Structure

Stacking (or stacked generalization) is an ensemble learning technique that combines multiple base classification models predictions into a new data set. This new data are treated as the input data for next layer classifier. Split the data set into training and testing sets and use the training set to train the model and testing set to test the model. Here, the main structure of the diagnosis, as shown in Fig.1, consists of two parts, Level 0(Base classification model) and Level 1(second layer classification model). There are four classifiers in level 0, Extra-Trees Classifier, Random Forest Classifier, AdaBoost Classifier, Grandient Boosting Classifier, SVM and XGB Classifier as meta-model.

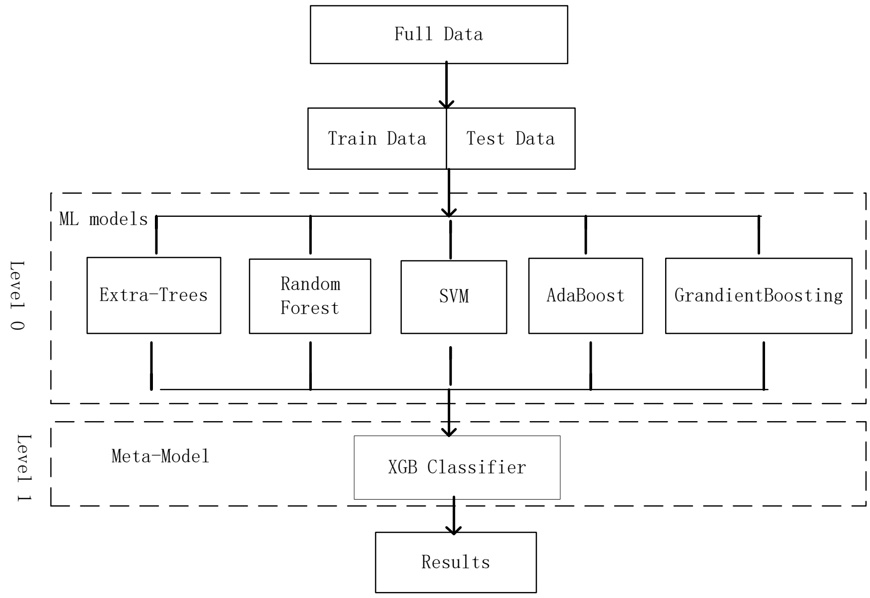

Fig.1 Stacking diagnosis model

3.2 Hyper-parameters optimize

Hyper-parameters are parameters that are not directly learnt within estimators. There are two generic approaches to parameter search: for given values and 1 exhaustively considers all parameter combinations. Inappropriate selection of hyperparameters will lead to problems of underfitting or overfitting. The initial parameters (e.g. n_estimators, Tree depth, forest size, C, min_samples_leaf, etc.) of the Classifier discussed above are optimized by cross-validated grid-search over a parameter grid.

3.3 Training for Base Classification models and meta-model

K-Fold CV is where a given data set is split into a K number of sections/folds where each fold is used as a testing set at some point. Here, use "k-fold cross-training" approach (similar to what is done in k-fold cross-validation) to train all the base classification models. The main steps of K-Fold CV for classifier described as follows:

(1) First the train data set is split into k folds. Select i fold as validation set. Train the weak classification model on the k-i folds and make predictions on the validation set.

(2) Use the validation set to calculate the mean square errorMSE_i

(3) Repeat the step(1) and step(2) for k times and calculate the average of the MSE

$$CV_k = \frac{1}{k}\sum_{i=1}^{k} MSE_i \tag{7}$$

As for the testing process, use the test data set to test the weak classification model at each step of k-fold training and average the test result as final result. The train and test process of each weak classification model is shown in Fig.2. By doing so for the five weak classification model, relevant predictions and test results for each model of our train data set are obtained and then all these predictions are concatenated to train the meta-model (XGB classifier).

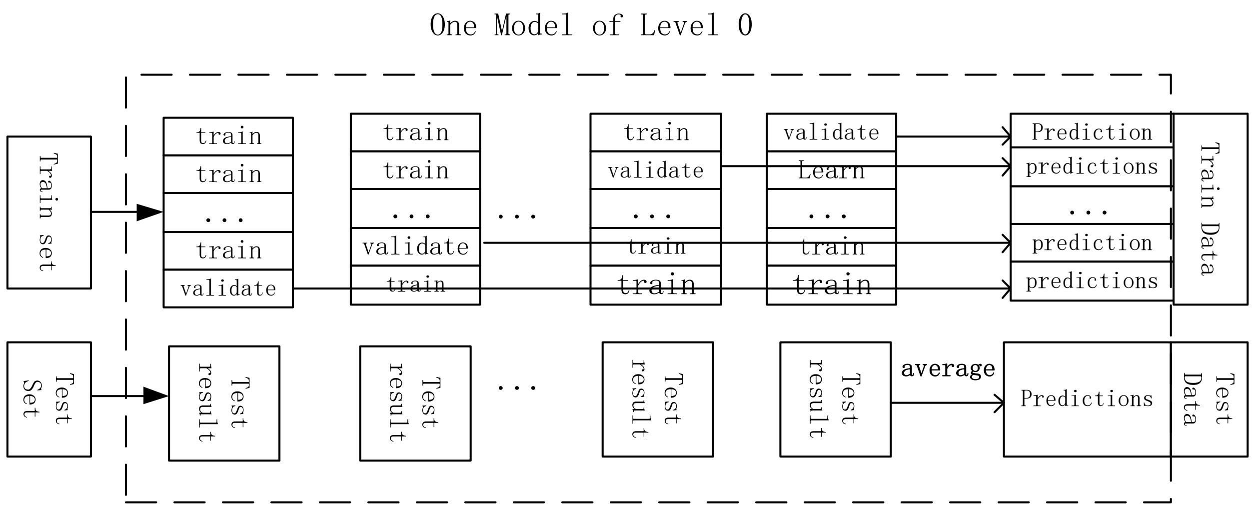

Fig.2 K-fold cross-training for base model

4. Experiment and Result Analysis

4.1 Experiment Result

In this paper, 300 sets of fault sample data are collected, which has been proved to be faulty and can reflect the transformer fault type. According to the proportion of 4:1, the sample data is randomly divided into two parts. And select the sample of the 240 groups as the training set, and the remaining 60 groups of samples as the test set. 6-fold cross-training" approach to train the classifiers of Level 0.Taking the content of H2, CH4, C2H6, C2H4 and C2H2 in oil as attribute information, and the transformer is divided into Six states including normal (N), low energy discharge (D1), high energy discharge (D2), Medium and low temperature overheating(T1) and High temperature overheating (T2). LabelEncoder technology is used to preprocess DGA sample data. As a result, T1,D1,N,T2,D2 are represented by 0,1,2,3,4 respectively. The Diagnosis accuracy of the proposed method reached 83.3%. Some the test results are shown in table 1.

Table.1 Test Results for DGA samples

No	Random Forest	Extra-Trees	AdaBoost	GrandierntBoost	SVM	Meta-model	Real status
1	4.0	3.0	3.0	4.0	3.0	3.0	T2
2	4.0	4.0	4.0	4.0	3.0	4.0	D2
3	3.0	3.0	4.0	3.0	3.0	3.0	T2
4	1.0	1.0	2.0	1.0	3.0	1.0	T1

4.2 Case Study

In a substation, the transformer SFSZ8-40000/110 produced in January 1998 and put into operation in March 1998. The transformer differential protection action caused the three side circuit breaker off, while the body heavy gas protection action [7]. DGA data is shown in table 2. The results of DGA analysis showed that acetylene content was 67 μL/L, and total alkyne content was 167.7 μ L/L, which exceeded the standard.

After the standardization process, the DGA data become (-0.27244779, -0.21453313, -0.04311869, -0.22668334, 0.47747329). The diagnosis results of the five models of Level 0 are 4, 4,1,4,3, respectively. The result of meta-model is 4, which means the fault type is high energy discharge fault.

Table 2. DGA data (μL/L)

Date	*CH4*	*C2H4*	*C2H6*	*C2H2*	*H2*	*CO*	*CO2*	*Total alkyne*
2007.11.24	0	54.43	17.17	0	8.24	905.54	5254.9	71.6
2008.06.25	27.25	60.5	12.97	67	59.2	1084	5513	167.7

5. Conclusion

Ensemble learning is a machine learning paradigm where multiple models are trained to solve the same problem and combined to get better results. By take advantage of the idea of stacking and five traditional classifiers, a new transformer fault diagnosis model was built and the diagnosis results show it is effective.

Acknowledgment

This work was supported by Overseas Study Project of Young and Middle-aged Key Teachers in Hebei Agricultural University.

References

[1] J. Dai, H. Song, G. Sheng and X. Jiang, "Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network," in IEEE Transactions on Dielectrics and Electrical Insulation, vol. 24, no. 5, pp. 2828-2835, Oct. 2017.

[2] Chaolong Zhang, Yigang He, Bolun Du, Lifen Yuan, Bing Li, Shanhe Jiang, Transformer fault diagnosis method

using IoT based monitoring system and ensemble machine learning, Future Generation Computer Systems, Volume 108, 2020,Pages 533-545
[3] J. Faiz and M. Soleimani, "Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis," in IEEE Transactions on Dielectrics and Electrical Insulation, vol. 25, no. 5, pp. 1798-1806, Oct. 2018.
[4] Xiaohui Yang, Wenkai Chen, Anyi Li, Chunsheng Yang, Zihao Xie, Huanyu Dong, BA-PNN-based methods for power transformer fault diagnosis,Advanced Engineering Informatics, Volume 39,2019,Pages 178-185.
[5] T. Kari et al., "An integrated method of ANFIS and Dempster-Shafer theory for fault diagnosis of power transformer," in IEEE Transactions on Dielectrics and Electrical Insulation, vol. 25, no. 1, pp. 360-371, Feb. 2018, doi: 10.1109/TDEI.2018.006746.
[6] Xue Wang, Tao Han."Transformer Fault Diagnosis Based on Stacking Ensemble Learning," in IEEJ Transactions on Electrical and Electronic Engineering,Volume15, Issue12, Pages 1734-1739,December 2020
[7] Y. Wang and L. Zhang, "Transformer fault diagnosis based on back-propagation neural network optimized by cuckoo search algorithm," 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE), 2017, pp. 383-386, doi: 10.1109/CCSSE.2017.8087962.
[8] Archana Chaudhary, Savita Kolhe, Raj Kamal,An improved random forest classifier for multi-class classification, Information Processing in Agriculture, Volume 3, Issue 4,2016,Pages 215-222.
[9] Chen, Siji et al. "A Strong Machine Learning Classifier and Decision Stumps Based Hybrid AdaBoost Classification Algorithm for Cognitive Radios." Sensors (Basel, Switzerland) vol. 19,23 5077. 20 Nov. 2019, doi: 10.3390/s19235077