Mu ZHANG, Kun LI, Huixin TIAN

Tianjin Polytechnic University

Multiple SVMs Modelling Method for Fault Diagnosis of Power Transformers

***Abstract**. For enhancing the accuracy of fault diagnosis for power transformers, a multiple SVMs scheme is proposed in this paper. In this scheme, SVM is used to establish the base classifier for its good performance and fast learning speed. Secondly, the several base classifiers based on single SVM will be combined by consulting ensemble techniques. And then a multiple SVM s method is obtained. The real gas records data from a power company is used to establish fault diagnosis system for power transformers based on the new multiple SVM s method. For comparison, the conventional methods are used to build fault diagnosis models by the same data. The experiments demonstrate the new multiple SVMs method has the best performance in both learning ability aspect and generalization ability aspect for fault diagnosis of power transformers.*

***Streszczenie.** Zaproponowano schemat SVM (support vector machine) w celu poprawy dokładności diagnostyki transformatorów mocy. W pprównaniu do metod konwencjonalnych proponowana metoda ma możliwość uczenia się i efektywnego wykorzystania bazy danych. (**Metoda wykorzystująca technikę SVM do diagnostyki transformatorów mocy**)*

Keywords: fault diagnosis, SVM, power transformers, ensemble technique.
Słowa kluczowe: diagnostyka transformatora, SVM.

Introduction

The power transformer plays an important role for the stable and reliable power transmission and distribution in modern power system. A failure of a power transformer may cause a serious outage of an electrical network. Therefore the power transformers must be examined strictly and periodically to find the incipient faults and keep away from the further deterioration as early as possible.

The researchers have done many efforts to detect the incipient faults of power transformers. And many detection methods have been proposed by different means. Dissolved gas analysis (DGA) method is one of the most widely used means. It is based on the fact that transformer faults, mainly in the form of thermal, arcing and partial discharge (PD), can be detected by analyzing dissolved gases in the insulation oil of a transformer [1-2]. The ratios of specific dissolved gas concentrations are processed using predefined criteria. These dissolved gasses contains mainly hydrogen (H2), methane (CH4), ethane (C2H6), ethylene (C2H4), acetylene (C2H2), carbon monoxide (CO), and carbon dioxide (CO2).

However, the above conventional methods do not fit the practical needs because the DGA methods can't identify the diagnosis comprehensively, especially when the diagnosis with complex, numerically rich and error contained relationship between dissolved gasses and fault types. Furthermore, the DGA methods are used to predefine the diagnosis criteria based on empirical studies, so different DGA methods often produce different judgment with the same dissolved gas record. With the fast development of artificial intelligence (AI), the AI methods are used in fault diagnosis of power transformers more and more for their efficient performance. The Artificial neural networks (ANNS) are used most widely to tackle the transformer fault diagnosis, because of their superior learning, generalization capabilities and fault-tolerate capabilities in practical applications [3-4]. Furthermore, Support vector machine (SVM) is a kind of method being suitable for the fault diagnosis. SVM has a good generalization capability even in the small sample cases of classification2). In recent years, SVM has been successfully applied for solving the classification problem. But there are still some problems in the practical application of SVM. The one is how to select the kernel function and the SVM parameters (the parameters of kernel function and regularization constant). It will make the performance of SVM unstable. Furthermore, the accuracy of single SVM is not satisfied for some practical applications.

In this paper the SVM will be used to establish the faulty diagnosis models. For the sake of enhancing the stability and the accuracy of the fault diagnosis system with single SVM, a SVM ensemble scheme is used to aggregate the single SVM models and establish a novel faulty diagnosis models for power transformers. The 760 real gas records data from a power company in china is used to build the fault diagnosis system based on multiple SVMs and test the performance of fault diagnosis system. The experiments demonstrate that the novel faulty diagnosis system can diagnose the faulty efficiently and the performance of fault diagnosis model based on multiple SVM is better than the one based on SVM.

The remainder of this paper is organized as follows. Section 2 describes the SVM briefly. Section 3 presents the multiple SVM method based on ensemble scheme. Section 4 shows the experiments for the fault diagnosis of power transformers in a power company by using the multiple SVMs. Conclusions from this study are summarized in section 5.

Support Vector Machine Review

Support vector machine (SVM) was originally introduced by Vapnik et al. 1999. Now SVM has become an increasingly popular technique for machine learning activities including classification, regression, outlier detection etc. It obtains real risk minimization by seeking for structural risk minimization. SVM integrates the optimal separating hyperplane with the kernel method. It could overcomes commendably such defects as dimensionality curse and overfitting that are apt to appear in some other conventional algorithms, neural net for instance. The idea of using SVM for separating two classes is to find support vectors to define the bounding planes, in which the margin between the both planes is maximized. The number of support vectors increases with the complexity of the problem [5-6].

To define SVM mathematically, the training data are first stacked into a $n \times m$ matrix X, where n is the number of observations and m is the number of variables. Denote x_i as a column vector representing the ith row of X. A $n \times n$ diagonal matrix Y with +1 and -1 entries is then used to specify the membership of each x_i in class +1 or -1. Observations in both classes satisfy.

(1) $$Y\left(Xw-e\gamma\right)\geq e$$

where e is a vector of ones. To determine w and γ, the following mathematical program is solved

(2) $$\min\frac{1}{2}\|w\|^2 \text{ s.t. } Y\left(Xw-e\gamma\right)\geq e$$

To determine w and γ, quadratic programming (QP) can be solved by finding the saddle point of the Lagrange functional (Scholkopf & Smola, 2002). The observations for which the corresponding Lagrange multipliers are positive are referred to as support vectors. With w and γ being solved, the class membership of x is determined as

(3) $$y_{pred}=sign\left(w^T x+\gamma\right)$$

For data that overlap, nonnegative slack variable ξ is introduced into the mathematical program

(4) $$\min\frac{1}{2}\|w\|^2+C\|\xi\| \text{ s.t. } Y\left(Xw-e\gamma\right)+\xi\geq e,\ \xi\geq 0$$

The first term in the objective function represents the model complexity and the second term represents the model accuracy (i.e. classification error in the training data). The regularization parameter C controls the trade-off between these two terms. Model underfitting occurs when C is too small.

The merit of support vector machine is that, by a kernel function, which is the inner product in the feature space, it ties to make the training data linear-separable in the high dimension feature space, thus achieve nonlinear-separable in the input space. Typical choices of kernel function include the follows:

Linear kernel function

(5) $$K\left(x_i,y_j\right)=x_i\cdot y_j$$

Polynomial kernel function:

(6) $$K\left(x_i,y_j\right)=\left[\left(x_i\cdot y_j\right)+r\right]^d$$

Radial-basis function (RBF) kernel (also named as Gaussian kernel)

(7) $$K\left(x_i,y_j\right)=\exp\left(-\lambda\left|x_i-y_j\right|^2\right)$$

Sigmoid kernel function:

(8) $$K\left(x_i,y_j\right)=\tanh\left(v\left(x_i^T\cdot y_j\right)-\delta\right)$$

where d is the exponent of polynomial function, r, λ, v and δ are kernel parameters.

Muliple Svms Scheme

Ensemble technique has become an increasingly popular technique for its great capability of enhancing the performance of individual machine. It combines the classifiers from multiple classifiers to produce a single classifier. The ensemble of machines is often called a committee machine. In a committee machine, an ensemble of classifiers is generated by means of a learning process; the overall classifiers of the committee machine are the combination of the individual committee members' classifiers. And various ensemble techniques have been investigated by many researchers. Several studies of ensemble techniques (such as boosting and bagging) in classification have demonstrated that these techniques are generally more accurate than the individual classifiers. To obtain a good ensemble, two main problems should be solved: how to construct accurate and diverse base classifiers and how to combine their outputs effectively. In a word, various ensemble techniques are different mainly in two ways: construct scheme and combining scheme [7].

Many methods for constructing ensembles have been developed. The most prominent ensemble machine generation methods are Bagging (Breiman, 1996) and Boosting (Freund & Schapir, 1996). Bagging employs bootstrap sampling to generate different training sets that consists of N train examples drawn randomly with replacement from the original training set of N items. Such training set is called bootstrap replicate. Bagging trains a classifier on each training set to build its constituent members, which can be generated in parallel.

Boosting works by repeatedly running a given weak learning machine on different distribution of training data and combining their outputs. In each iteration, the distributions of training examples depend on the performance of the machine in the previous iteration. The main idea of boosting is to train a sequence of classifier with the so-called boosting sets so that each subsequent classifier concentrates mostly on the errors made by the previous ones. This is achieved by maintaining a set of weights over the whole training examples.

After constructing a good ensemble of classifiers, it is also necessary manipulate the output values that are given to the learning algorithm. A variety of combining scheme have been developed and used for classifier ensemble to manipulate the output in the literatures. The most often used and popular schemes are the average, weight and majority vote combining. Average approach averages the individual classifier outputs across all the members. Weight approach combines the outputs according to the performance of individual classifier. That is to say, the better classifier will be given a big value of weight, otherwise will be given a small value of weight. In majority vote approach the output will be controlled by the classifiers that have majority votes. It is the most popular aggregation method for its easy implementation.

Ensemble technique has become an increasingly popular technique for its great capability of enhancing the performance of individual machine. It combines the classifiers from multiple classifiers to produce a single classifier. The ensemble of machines is often called a committee machine. In a committee machine, an ensemble of classifiers is generated by means of a learning process; the overall classifiers of the committee machine are the combination of the individual committee members' classifiers. And various ensemble techniques have been investigated by many researchers. Several studies of ensemble techniques (such as boosting and bagging) in classification have demonstrated that these techniques are generally more accurate than the individual classifiers. To obtain a good ensemble, two main problems should be solved: how to construct accurate and diverse base classifiers; how to combine their outputs effectively [8].

In order to improve the performance of fault diagnosis system, a multiple SVMs scheme is used to establish the fault diagnosis model. Firstly, the individual SVMs are constructed with different kernel function and different kernel parameters. The kernel functions involve Polynomial kernel, RBF kernel and Sigmoid kernel. The kernel parameters and regularization constant C are chosen randomly in their valid range. Therefore, different Polynomial kernel SVM classifiers, RBF kernel SVM classifiers and Sigmoid kernel SVM classifiers are established by this approach. Secondly, the SVM ensemble is established using bagging ensemble technique. The data sub-sets with same number are obtained by bootstrap method from original data set. The individual SVMs are chosen randomly in above generated SVM classifiers as weak learning machine of Bagging. Every individual SVM is trained by corresponding data sub-set. At the end, the weak learning machines can be aggregated by weight vote combining scheme. So a fault diagnosis model based on SVM ensemble is established.

Experiments

The 760 real gas records data from a power company in china is used to build the fault diagnosis system for the 750kv oil-immersed power transformers based on multiple ELMs and test the performance of fault diagnosis system. The fault types are confirmed by the transformer diagnosis experts according to the IEEE criteria for the interpretation of DGA methods and other related information. Table 1 specifies the five types of transformer faults which are contained in the 760 gas data.

Firstly, the single ELM classifiers are established by above 760 gas data. The inputs of SVM are the concentration of five gases: H2, CH4, C2H6, C2H4, and C2H2 which were used in DGA by IEEE criteria. And then the SVM classifiers are combined to establish the multiple ELM fault diagnosis system for power transformers by the ensemble scheme presented in the section 3.

For comparison, another conventional empirical method traditional ANN method and single SVM method are conducted the experiments. The IEEE standard DGA method is applied for fault diagnosis based on the same gas data. The most widely used BP neural network is selected as a traditional ANN method to establish the fault diagnosis system. The input and output nodes in BP network are the same as the ones in SVM. The number of hidden nodes of BP network is determined by experiments. And the training data are also the same gas data. The performance of above three fault diagnosis models are compared in two ways: learning ability based on training data, and generalization ability based on testing data.

Table 1. The five types of power transformer faults

fault code	fault type	data
0	No fault	558
1	<300℃ thermal fault	53
2	300℃-700℃ thermal fault	51
3	>700℃ thermal fault	47
4	low-energy partial discharge	51
-	total	760

Table 2 Learning ability of the model based on different methods by 760 gas data

Method for fault diagnosis	Fault diagnosis accuracy (%)
IEEE standard	84.82
BP neural network	90.26
Single SVM	90.35
Multiple SVMs	92.16

Table 2 shows the comparison of learning ability between multiple ELMs model and the model based on other methods (involving DGA, BP network, and single ELM) using the same 760 real gas records data. The results demonstrate that the performance of the model based on artificial neural network methods is better than the one based on conventional empirical methods (DGA). Among the single ANN methods, the learning ability of ELM is better than traditional BP neural network. The proposed multiple ELMs method has the best performance of learning for fault diagnosis of power transformers.

As shown in Table 3, the fault diagnosis accuracy by test data is presented. The generalization performance of BP neural network model is the worst. And the fault diagnosis accuracy by testing data even is not better than the one by training data. The learning accuracy and the generalization accuracy of ELM approximate. The single ELM has better generalization performance. However when the single ELMs were combined, the generalization ability is improved significantly. Obviously, the multiple ELMs model has the best accuracy among all methods by the testing data.

Table 3 Generalization of different methods by testing data

Method for fault diagnosis	Fault diagnosis accuracy (%)
BP neural network	89.61
Single ELM	91.89
Multiple ELMs	94.26

Conclusion

Aiming at enhancing the accuracy of fault diagnosis for power transformers and overcoming the limitations of conventional empirical methods, a new multiple SVMs scheme is proposed to establish the fault diagnosis system. In the new scheme, SVM is selected as the basic intelligent method to establish the system for its better performance when it is used to solve the faulty classification problem. The 760 real gas records data from a power company in china is used to build the fault diagnosis system for the 750kv oil-immersed power transformers by the new multiple SVMs method. For comparison, the IEEE standard, BP neural network and single SVM are also used for diagnosing the power transformers fault. The learning ability and the generalization ability of different methods are test by training data and test data. The results of experiments demonstrate that the fault diagnosis system based on new multiple SVMs method has the best performance in both aspects.

REFERENCES

[1] IEEE, “IEEE Guide for the Interpretation of Gases Generated in Oil Immersed Transformers, standard ed.” Picataway, NJ: IEEE Press, 1992.

[2] “Interpretation of the Analysis of Gases in Transformers and Other Oil-Filled Electrical Equipment in Service:” IEC Publication 599, 1978.

[3] Y. C. Huang, “Evolving Neural Nets for Fault Diagnosis of Power Transformers.” IEEE Transactions on Power Delivery, 18, No. 3, 843-848.

[4] Y. C. Liang, X. Y. Sun, D. H. Liu, “Application of Combinatorial Probabilistic Neural Network in Fault Diagnosis of Power Transformer.” Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22(2007), 1115-1119.

[5] Vapnik, V. N. (1999). The nature of statistical learning theory. New York: Springer.

[6] Mayoraz, E., & Alpaydin, E. (1999). Support vector machines for multi-class classification. Proceedings of International Workshop on Artificial Neural Networks, 2, 833–842.

[7] G. I. Webb, “NultiBoosting: A Technique for Combining Boosting and Wagging,” Machine Learning, 40(2000), No. 2, 159-196.

[8] Y. Freund, R. Schapire, “Experiment with a new boosting algorithm,” Proc. Of the 13th Int. Conf. on Comput. Learning Theory, San Francisco, Morgan Kaufmann, 1996, 208-219.

***Authors**: prof. huixin TIAN, Politechnika Poznańska, Tianjin Polytechnic University, E-mail: icedewl@163.com.*