Fault Diagnosis Model of Transformers Based on Neighborhood Rough Set and Relevance Vector Machine

Chen Jialin[1], Zhang Mingyu[2]*, Duan Jiahua[1]

1Yunnan Province Energy Investment Group co., LTD, Yunnan Kunming, 650021;

2 Energy Industry Development Institute of Yunnan Province Energy Investment Group co., LTD, Yunnan Kunming, 650021)

Received 12 September 2014, www.cmnt.lv

Abstract

Transformers are important devices in power supply system. Since the testing data in fault diagnosis of transformers have features such as high dimension, uncertainty and incompleteness, existing fault diagnosis methods are incapable of dealing with the high-dimensional and abnormal data, and the diagnosis precision needs further improvement. This paper innovatively proposes a fault diagnosis model for transformers based on neighborhood rough set (NRS) and relevance vector machine (RVM). It integrates advantages of NRS in handling uncertain and incomplete information and RVM's capability in dealing with high-dimensional data. Based on data provided by Yunnan Kunming Power Supply Company in fault diagnosis of transformers, case study has shown that the testing diagnosis rate of four classifiers in this model are 92.55%, 93.87%, 93.44% and 92.28% respectively, higher than the diagnosis precision of RVM, which indicates this model enjoys better diagnostic capability.

Keywords: neighborhood rough set; relevance vector machine; transformers; fault diagnosis

1 Introduction

The operating status of transformers directly affects the safety level of electrical system. Therefore, it is vital to establish a fault diagnosis model based on online supervision of transformers' operating status to ensure the safe operation of power system [1].

Currently, dissolved gas analysis (DGA) [2] is a common method in the fault diagnosis of oil-immersed transformers. Traditional methods like three-ratio method, Rogers method and Dornerburg method recommended by International Electrotechnical Commission (IEC) have disadvantages of coding loss and absolute coding limits [1-2]; artificial neural network (ANN) [3] of artificial intelligence algorithms has limitations of over-fitting, which will lead to local optimum; Bayesian networks (BN) [4] requires a large amount of sample data; support vector machine (SVM) [5] has its difficulty in determining the rule coefficients and meeting Mercer conditions of kernel function. As a machine learning algorithm of sparse Bayesian learning theory, relevance vector machine (RVM) [6] combines theories of Bayesian, Markov, MLE and auto-correlated a priori determination. It not only keeps the advantages of SVM, but also overcomes the shortcomings: it gets rid of the restriction of Mercer conditions; most weights of basic functions are zero, which is more sparse than SVM; there is no need to ruled coefficients and the parameters are acquired without the use of cross validation; the output of posterior probability distribution function [7] is available.

It has made some progress in the area of RVM's application in transformers' fault diagnosis in recent tears. Case study has shown that it is more advantageous than SVM [8]. However, if we fail to exclude abnormal data and acquire the sensitive features as the input of classified models from the large amount of original chromatographic characteristics of dissolved gases in the oil, and adopts the input of original features, it will make the calculation more complex and the classification outcome is not that satisfactory. Rough set theory (RST) [9] has gained wide application in terms of characteristics selection, knowledge exploration and rules extraction [10]. But traditional RST cannot directly deal with numerical variables, which poses inconvenience to the wide use of RST [11]. A new algorithm based on local optimization and neighborhood rough set feature extraction, offsets the weakness of traditional RST and can effectively extract the sensitive features [12-14].

From this perspective, this paper proposes a fault diagnosis model of transformers based on neighborhood rough set and relevance vector machine (short for NRS-RVM model), over -comes the shortcomings of single relevance vector machine in treating the failure of transformers, and integrates the abilities of neighborhood rough set on handling incomplete information and the fastness of extracting sensitive features, which further improve the binary classification accuracy of relevance vector machine.

2 Related Principles

2.1 FEATURE SELECTION ALGORITHM BASED ON ROUGH SET THEORY

Let the information system IS=<U, A, V, f> be a decision table. If A=C∪D={a1, a2, …,am} is an attribute set, the intersection between condition attribute set C and decision attribute set D is an empty set, V is the value range, U={x1, x2, …,xn} is the domain, and f: U×A→V is the information function. For every xi∈U, condition attribute subset B$\subseteq$C, the neighbor -hood of xi is defined as $\delta_B\ (x_i)$={xj | xj∈U, Δ_B (xi, xj) $\le \delta$ }, where Δ refers to distance function expressed by p norm usually, and δ is the neighborhood radius.

***Corresponding author** e-mail: hanjiesmile@163.com

Let NDT =<U, C∪D, N, f>, $\forall$ B $\subseteq$ C be a neighbourhood-based decision-making system. D divides U into N object subsets X1, X2, …, XN deciding 1, 2, …, N respectively. Define the lower approximation of D as NBD= $\bigcup_{j=1}^{N} \underline{N_B} X_j s$, where NBXj={xi | δ_B (xi) $\subseteq$ Xj, xi∈U}, and δB (xi) is the neighborhood information granule generated by B. The lower approximation of D (also known as the decision positive domain) is POSB(D). The value of decision positive domain reflects the separable degrees in the given attribute set. The larger the domain is, the less the overlapping section of each type (i.e. boundary) is. Therefore, define the dependence of D on B as γB(D)=|POSB(D)|/ |U|, where |·| is the cardinal number of set.

2.2 RVM-BASED CLASSIFICATION MODEL

Let x1, x2, …, xN be the input vector of N, t= (t1, t2, …,tn)' the target vector, target value ti =0 or, ti =1 $\mathbf{w} = (w_1, w_2, \cdots, w_n)'$ the adjustable weight vector, $K(\mathbf{x}, \mathbf{x}_i)$ the kernel function, and x is the input, the output of classification model is $y(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{n} w_i K(\mathbf{x}, \mathbf{x}_i) + w_0$. Define the priori probability distribution density function of every weight wi as p(wi|ai)= $(\frac{a_i}{2\pi})^{\frac{1}{2}} \exp(-\frac{1}{2} a_i w_i^2)$, where ai is the hyper- parameter deciding the priori distribution of weight wi. The priori probability distribution density function of weight vector $\mathbf{w}$ is p(w|a) = $\prod_{i=0}^{N} f(w_i | 0, \alpha_i^{-1})$, where a=(a1, a2, …, an)' is the $n+1$ dimensional hyper parameter vector deciding w. f(·) is the distribution density function of normal distribution N(0, ai-1).

Suppose the target value of training sample set {xi, ti} $_0^N$ is independent, and the noise of data matches the Gaussian distribution N(0, σ^2), the likelihood function of training sample set is: p(t|w, σ^2)= $(2\pi\sigma^2)^{-\frac{N}{2}} \exp(-\frac{\|\mathbf{t} - \mathbf{\Phi w}\|^2}{2\sigma^2})$. Where Φ(xi)= [1, K(xi,x1), K(xi,x2), … , K(xi,xN)]', Φ [Φ(x1), Φ(x2), …, Φ(xN)]'. In binary classify- cation, input x is given and the output y(x, w) is included in t-class affiliated posterior probability p(t|w). Generalize the linear model by using Logistic Sigmoid function on the output y(x, w). Suppose p(t|x) obeys Bernoulli distribution, the likelihood function of training sample set is: $p\{t|\mathbf{w}\} = \prod_{i=1}^{N} \sigma[y(\mathbf{x}_i, \mathbf{w})]^{t_i} \{1 - \sigma[y(\mathbf{x}_i, \mathbf{w})]\}^{1-t_i}$, where σ(·) is the Logistic Sigmoid function.

Let $\mathbf{x}_*$ be the new input, and t_* the corresponding target value, p(w|t, a) is priori probability distribution density function of weight vector. p(a|t) is the marginal likelihood function and the probability prediction is: $p(\mathbf{t}_* | \mathbf{t}) = \int p(\mathbf{t}_* | \mathbf{w}, \boldsymbol{\alpha}) p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}) p(\boldsymbol{\alpha} | \mathbf{t}) d\mathbf{w} d\boldsymbol{\alpha}$.

Since the priori probability distribution density function of weight vector p(w|t, a) and marginal likelihood function p(a|t) cannot be decided by integrals, the approximation of Laplace-based method is adopted shown as follows.

(1)a=(a1, a2, …, aN)', the original hyper parameter vector and posterior distribution are given. The method of maximum weight vector posterior probability distribution function p(w|t, a) is adopted to estimate the "most probable" weight vector wMP. Since p(w|t, a) $\propto$ p(t|w) p(w|a), it is equivalent to the maximum of a regularized Logistic logarithm likelihood function:

$$\log[p(\mathbf{t}|\mathbf{w}) p(\mathbf{w}|\boldsymbol{\alpha})] = \sum_{t=1}^{N} \{t_i \log \sigma[y(\mathbf{x}_i, \mathbf{w})] +$$

$$(1 - t_i) \log\{1 - \sigma[y(\mathbf{x}_i, \mathbf{w})]\}\} - \frac{1}{2} \mathbf{w}' A \mathbf{w},$$

where A=diag(a0, a1, …, an) is the diagonal matrix. Secondary Newton algorithm is used to acquire the maximum.

1) Obtain the gradient vector of w based on the above equation: $\mathbf{g} = \nabla_{\mathbf{w}} \log[P(\mathbf{t}|\mathbf{w}) p(\mathbf{w}|\boldsymbol{\alpha})] = \mathbf{\Phi}'(\mathbf{t} - Y) - A\mathbf{w}$, where B = diag(β1, β2, …, βN), βi = Yi (1- Yi), Yi=(Y1, Y2 , …, Yn)' and Yi= σ[y(xi, w)].

2) Obtain the Hessian matrix of w:

$$H = \nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \log[P(\mathbf{t}|\mathbf{w}) p(\mathbf{w}|\boldsymbol{\alpha})] = -(\mathbf{\Phi}' B \mathbf{\Phi} + A)$$

3) Acquire wMP by iterative weighted least square method: ∆w=-H-1g, wMP=wMP +∆w.

(2) The posterior probability distribution density function of Gaussian approximation weighted vector is p(w|t, a)≈f(wMP, ∑), where wMP is the posterior mean vector of weighted vector, and∑=(-H|wMP)-1 is the posterior covariance matrix of weighted vector.

(3) Mackay[15] is adopted to maximize the marginal likelihood function p(a|t) to update the hyper parameter vector a, i.e., ainew=γi/w2MPi, γi=1-ai∑ii, where Σ_{ii} is the diagonal entry of i in∑, and wMPi is the i component of wMP.

Repeat the above procedures until the condition of convergence is met. In practice, most ai is close to infinity, and its corresponding weight is wi =0; others are tend to finite values. The set of learning sample xi corresponding with the non-zero wi is called as relevance vector (RV).

3 Establishment of transformer' fault diagnosis model

An information system of fault diagnosis in transformers is given. In the training process of single relevance vector machine itself, space complexity of matrix operations arrives at o(N2), the total time complexity reaches to o(N3), redundant information and conflict object affect the generalization of RVM, which will lead to the decrease of classify-cation diagnosis rate of transformers' fault diagnosis system. Therefore, attribute reduction is necessary. Meanwhile, neglecting the effects of different attributes on the fault diagnosis will result in invalid and incorrect outcomes. So, training sample set after reduction needs further treatment of attribute weighting. This paper adopts neighborhood rough set to offset the shortcomings of RVM's application in transformers' fault diagnosis[16].

Firstly, rapid reduction algorithm is used to reduce attributes. Second, based on condition attribute's dependence on decision attribute, each attribute is weighted. Next, feature vector set formed after attribute reduction and numerization is input into RVM as the training sample. Finally, test the training outcomes with the testing sample set and output the

results of classification, detailed procedures are shown in Figure 1 and detailed calculation is shown as follows[17].
Input: decision table 〈U, C∪D, N, f〉
Output: classification of transformer failure
Step 1: Complete the training samples and samples to be classified and make them discrete.
Step 2: If it takes training task, start from step 3; otherwise, start from step 6.
Step 3: Attribute reduction
Step 3.1: Initialize the feature subset and sample set being tested, i.e. let red=φ, sample = U.
Step 3.2:
flag=1;
while sample=ϕ
for each ki =C\red
NDTi = 〈U, red∪ki∪D, V, f〉
Initialize POSi=ϕ;
for each aj∈sample
calculate the neighborhood δ(aj) of aj under NDTi ;
if each sample's decision attribute D of δ(aj) shares the same value
POSi= POSi∪aj
end if
if flag=1
γi= POSi/sample
end for
flag=0
find out maximum POSi and corresponding ki;
if POSi ≠ ϕ
red=red∪ki
sample= sample/ POSi
else
quit while cycle
end if
end while
Step 3.3: return red
Step 4: Attribute weighting, i.e., the training sample set after reduction is multiplied with corresponding importance γi.
Step 5: Train the classifiers of RVM and acquire the weight vector w=(w1, w2, …, wn).
Step 5.1: Initialize hyper parameter vector a=(a1, a2, …, an).
Step 5.2: Keep the hyper parameter vector stable and give its posterior distribution form; secondary Newton method is used to acquire wMP.
Step 5.3: make quadratic approximation of logarithm posterior probability distribution density function around its peak value based on Laplace method and obtain the Gaussian approximation covariance matrix of weight posterior probability distribution density function at wMP: ∑=(-H|wMP)-1.
Step 5.4: a is given, and recalculate wMP.
Step 5.5: Recalculate hyper parameter ainew = γi / w2MPi, where γi=1- ai∑ii.
Step 5.6: If it is converged, turn to Step 6; otherwise, redo it from Step 5.3.
Step 6: As with the new failure of transformer, calculate $y(\mathbf{x}_*, \mathbf{w}) = \sum_{i=1}^{N} w_i K(\mathbf{x}_i, \mathbf{x}_*) \ \ + w_0$, and obtain the category of fault.

4 Results and analysis of simulation experiments

Transformer failure data collected by Yunnan Kunming Power Supply Company has 292 groups in total. Take 200 groups as the training set and the rest 92 as the test set. Binary classification is adopted respectively to transform multi-classification issue into many binary classification issue to construct a RVM model based on single relevance vector machine and neighborhood rough set. The volume fraction of hydrogen, methane, ethane, ethylene and acetylene acts as the input. In order to keep the definition domain of kernel function K(x,xi) within [-1,1], the five kinds of gas is treated as follows. xi =mi /max (mi), i=1, 2, …,5, where mi is the volume fraction of the five gases. Four classification instruments are used to define five states of transformers as normal state, discharging with low energy, discharging with high temperature, overheated state in medium and low temperature and overheated state in high temperature. If the output value of classification instrument is 1, the transformer is in failure; if the output value of classification instrument is -1, the transformer is in another failure. Kernel functions of two methods adopt the radial basis function (RBF) K(x, xi)=exp(-||x- xi||2/ 2γ2, and spread factor γ of kernel functions are both 0.5. The fault diagnosis model of transformers is shown in the figure 1.

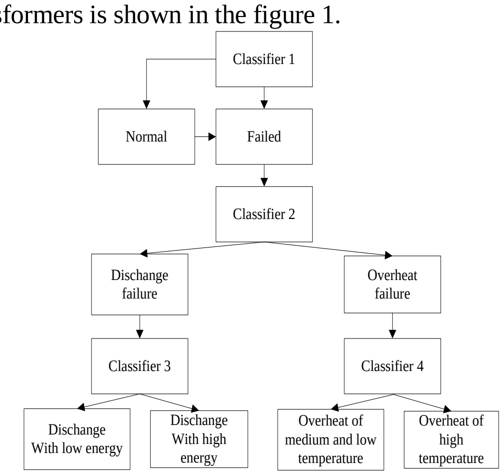

FIGURE 1 The fault diagnosis model of transformers

The comparison between two diagnosis method is presented in table 1. It indicates that the fault diagnosis model based on RVM and neighborhood rough set enjoys higher accuracy. To verify whether the prediction accuracy is prominent, this paper has conducted Wilcoxon's Sign Rank Test. Table 2 has proved that this model is qualified, which reflects this method effectively improves the prediction accuracy of single RVM model.

TABLE 1 Comparison between two fault diagnosis models of transformers

Model	Classifier	Training time /s	NV	Test diagnosis rate /%
RVM	1	0.5872	11	88.99
	2	0.5913	9	90.49
	3	0.7196	10	91.33
	4	0.1028	9	89.67
NRS-RVM	1	0.6764	10	92.55
	2	0.6722	8	93.87
	3	0.5982	10	93.44
	4	0.1931	7	92.28

*NV: The number of vectors

TABLE 2 The result of Wilcoxon's sign rank test

Model	N	Rank	Sum	P-Value
(RVM)-(NRS-RVM)	63	44.54	1623	0.000

*N: The average absolute error of this model is smaller than the number of single RVM models.

5 Conclusion

Causes of transformers failure are complex, which cannot be depicted compre- hendsively and accurately by a single method. This paper proposes a new fault diagnosis method based on neighborhood rough set (NRS) and relevance vector machine (RVM) and it (1) effectively excludes the abnormal data, which improve the sensitivity of RVM; (2) has carried out attribute reduction and weighting based on the determination of attribute importance. It takes full con- sideration of the advantages of RVM, and offsets the disadvantages of RVM via neighborhood rough set as well, which makes the mixed algorithm more robust and easier to promote. Results of simulation experiment show that the test diagnosis rates of four classifiers are 92.55%, 93.87%, 93.44% and 92.28% respectively, which is higher than the ones in the model based on RVM; this method is effective and enjoys high accuracy of diagnosis.

References

[1] Zhu D H, Yan Z, Tan K X, et al. Present Situation and Prospects of Condition Monitoring and Fault Diagnosis Technology for Electrical Equipments. Beijing: Electrical Equipment, 2009: 244-302.
[2] Tang W H, Wu Q H. Condition monitoring and assessment of power transformers using computa- tional intelligence. New York, USA: Springer -Verlag Press, 2011: 95-104.
[3] Sun J X. Modern Pattern Recognition. Beijing: Higher Education Press, 2008: 454-494.
[4] Zhao W Q, Zhu Y L, Wang X H. Combinatorial Bayes network in fault diagnosis of power transformer . Electric Power Automation Equipment, 2009, 29(11): 6-9.
[5] Zhang Z, Zhao W Q, Zhu Y L. Power transformer condition evaluation based on support vector regression. Electric Power Automation Equipment, 2010, 30(4):81-84.
[6] Tipping M E. The relevance vector machine. Advances in Neural Information Processing Systems, 2001, 12: 652-658.
[7] Tipping M E. Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research, 2001, 1(3): 211-244.
[8] Yin J L, Zhu Y L, Yu G Q. Relevance vector machine and its application in transformer fault diagnosis. Electric Power Automation Equipment, 2012, 32(8): 130-134.
[9] Vapnik V N. The nature of statistical learning theory. New York, USA: Springer-Verlag Press, 1995: 181-218.
[10] Pawlak Z. Rough sets. International of Computer and Information Science, 1982, 11(5): 341-356.
[11] Zhai Jinxiu, Zhang Zhousuo, He Zhengjia. Fault diagnosis model based on multi-wavelet packet and neighborhood rough set. Journal of Vibration, Measurement & Diagnosis, 2013, 33(S1):137-140.
[12] Hu Qinghua, Liu Jinfu, Yu Daren. Mixed feature selection based on granulation and approximation. Knowledge-based Systems, 2008, 21(4): 294-304.
[13] Hu Q H, Yu D R, Xie Z X. Neighborhood classifiers. Expert System with Applications, 2008, 34(2): 866-876.
[14] Hu Q H, Yu D R, Liu J F, et al. Neighborhood rough set based heterogeneous feature subset selection. Information Sciences, 2008, 178(18): 3577-3594.
[15] Mackay J C. The evidence framework applied to classification networks. Neural Computation, 1992, 4(5): 720-736.
[16] Hong Zhiyong, Liu Hua, Deng Weibin, et al. One Novel Text Classification Approach Based on Rough Set and RVM. Computer Simulation, 2010, 27(7): 183-186.
[17] Han Hu, Dang Jianwu, Ren En'en. Research of Support Vector Classifier Based on Neighborhood Rough Set. Computer Science, 2010, 37(2): 229-285.

Authors

Chen Jialin, born in February, 1973, Wenzhou District, Zhejiang city, P.R. China

Current position, grades: department head, Engineer China of Yunnan Province Energy Investment Group co.
University studies: B.Sc. and M.Sc. from Kunming University of Science and Technology in China.
Scientific interest: energy economy, electric Power Automation.
Publications: more than 10 papers published in various journals.
Experience: 16 years of experience in Power Resource Management, 5 scientific research projects.

Zhang Mingyu, born in May, 1980, Xuzhou District, Jiangsu city, P.R. China

Current position, grades: Vice President China of Yunnan Province Energy Investment Group co.
University studies: M.Sc and D.E from Kunming University of Science and Technology in China.
Scientific interest: Energy Industry, electric Power Automation
Publications: more than 20 papers published in various journals.
Experience: 5 years of experience in Power Resource Management, 10 scientific research projects.

Duan Jiahua, born in May, 1981, Baoshan District, Yunnan city, P.R. China

Current position, grades: Technical Supervisor, China of YPIC Renewable Energy Development CO., LTD.
University studies: Undergraduate course from Kunming University of Science and Technology in China.
Scientific interest: energy economy, electric Power Automation.
Publications: more than 10 papers published in various journals.
Experience: 5 years of experience in Power Resource Management, 5 scientific research projects.