Proc. of the 5th WSEAS/IASME Int. Conf. on Electric Power Systems, High Voltages, Electric Machines, Tenerife, Spain, December 16-18, 2005 (pp1-5)

Genetic Programming Based Polynomial Networks Model for Insulation Fault Diagnosis of Power Transformers

Zheng Zhang[1], Dengming Xiao[1] and Yilu Liu[2]
[1]School of Electronic Information and Electric Engineering, Shanghai Jiao Tong University, 1954 Road Huashan, Shanghai, CHINA
[2]Department of Electrical & Computer Engineering, Virginia Polytechnic Institute and State University,
302 Whittemore, Blacksburg, VA 24061, USA

Abstract: A Genetic Programming based Polynomial Networks Model (GPPNM) is presented in this paper to promote the diagnostic performance of incipient insulation fault of power transformers. Other than conventional hierarchical architecture to build polynomial networks, the proposed GPPNM constructs it using tree-like structure of Genetic Programming (GP). By means of flexible selection of low-order polynomial functions and feature variables in each node of structure, the polynomial networks is evolving in the global search space by generations to capture the complex and numerical knowledge relationships between dissolved gases and fault types. The proposed model has been applied on the actual fault records and compared with conventional method, artificial neural networks method and self-organizing polynomial networks (SOPN) method. The numeric test testifies that the GPPNM requires less prior knowledge in the process of construction of diagnosis model and has better performance than other methods.

Key-Words: - Power transformer, Insulation fault diagnosis, Dissolved gas analysis, Polynomial networks, Group Method of Data Handling, Self-organizing polynomial networks, Genetic programming

1 Introduction

Oil-immerged power transformers are the most pivotal devices in the power delivery system. The insulation condition of these devices has a great influence on the stability of power system. Therefore, it is essential to develop fault diagnosis system that can identify insulation fault types inside the power transformers to provide reliable maintenance information before they deteriorate to a severe state.

In the techniques of insulation diagnosis, dissolved gas analysis (DGA) based methods have been proved most effective to distinguish insulation fault types by means of concentration of special dissolved gases of power transformers. When some kind of insulation fault occurs, insulation oil decomposes into a series of characteristic gases under the condition of electrical and thermal stress. Conventional DGA methods, such as the key gases analysis, the Donrnenburg method, Rogers' gas ratio method and IEC/IEEE standard criteria, are developed based on experience of experts and statistic results to interpret the relationship between dissolved gases and insulation fault types [1-4].

Fuzzy theory based diagnosis systems are developed to improve the vagueness of boundary among criteria [5]. However, it is difficult to conclude more fuzzy rules in the case of multidimensional input when better performance is required. Artificial neural networks (ANNs) based methods are also introduced to diagnose the fault types of power transformers [6-7]. But, the choosing of neuron type and networks structure is mainly on the prior knowledge of authors and trial-and-test. In the meanwhile, the convergence of network parameters learning algorithm also have a great influence on the performance of ANNs. As another method of nonlinear system modeling, Group Method of Data Handling (GMDH) algorithm was introduced by Ivakhnenko in the early 1970's to model complex nonlinear system [8]. The main characteristic of GMDH is that it is a self-organizing and provides an automated selection of essential input feature variables without prior information on the relationship among input-output variables. Self-organizing Polynomial networks (SOPN) is a useful GMDH-type algorithm which has a hierarchical architecture [9]. The output of the each node of layers in SOPN structure is obtained using several types of high-order polynomial such as linear, quadratic and modified quadratic of input variables. These polynomials are called partial descriptions (PDs). Although the SOPN is structured by a systematic design procedure, it has some drawbacks to be solved. The selection principle and numbers of good PDs in each layer are still chosen in advance and discarded PDs in former layers can not be reused by later layers. Moreover, the SOPN algorithm is a heuristic method so it does not guarantee that the obtained SOPN is the best one for nonlinear system modeling. Then further study is necessary to promote the performance of SOPN based fault diagnosis of power transformers [10].

Recently, Genetic Programming (GP) algorithm has received considerable attention in the structure optimization domain and obtained many successful applications in the area of discovering nonlinear relationship of input-output system [11]. In this paper we present a new diagnosis system for insulation fault of power transformer using Genetic Programming based Polynomial Networks Model (GPPNM) in order to alleviate the above-mentioned drawbacks of the SOPN and promote the diagnosis performance.

This paper is organized as follows. An overview of GP algorithm is described in Section 2. The design methodology of GPPNM and flowchart of GPPNM based diagnosis system are described in Section 3. A hierarchical classification strategy for insulation fault types is shown in Section 4. Numeric test results of the proposed GPPNM-based diagnosis system are presented in Section 5. Finally, Section 6 draws the conclusion.

2 Overview Of GP

Prof. John Koza introduced the concept of GP in his creative research work. GP extends the chromosome of Genetic Algorithm (GA) into a combination of special programs to construct alternative solution to the problem. The individual in the population of GP is represented in the form of tree-like nodes or programs. Two kinds of nodes- function and terminal nodes- are used in the structure. The function nodes act as a program to fulfill a special task. Arithmetic operators, mathematical functions, boolean operators, conditional operators are usually selected in the function sets of GP. User-defined functions and automatic define functions also can be used in GP to solve special task. The terminal nodes stand for the basis unit of problem. Special constants, random numbers and input attributes are usually used in the terminal sets. The choice of function sets and terminal sets vary on the problem to be solved.

The population of GP evolves using the Darwinian principle of survival of the fittest. GP begins with a population of randomly created programs using some kinds of tree-grown algorithms. They are possible solutions to a given problem. In every generation, the fitness of each individual is evaluated. For the next generation, the survive possibility of individual is based on its fitness. The population evolves over a number of generations through the operation of various operators, such as reproduction, crossover and mutation. The dynamic tree-like structure of individual ensures the global search capability of GP to find proper structure and parameters of solution. The final result of GP (the best solution found) is the fittest solution produced along all generations when stopping rule is satisfied.

3 Genetic Programming Based Polynomial Networks Model (GPPNM)

Other than the hierarchical architecture of conventional polynomial networks, the GPPNM utilize tree-like structure of GP to construct the networks. Then evolutionary search strategy is adopted to learn the best structure which describes the nonlinear relationship of between system input and system output.

3.1 Representation of individual

A high-order polynomial can be constructed by a set of low-order polynomials and variables. In the function set of GPPNM, we take 16 different second-order polynomial functions that take two arguments [12]. Using the tree-like structure, it is very easy to create various high order polynomials. Table 1 lists all these second-order polynomial functions. The terminal set contains feature variables of system.

Table 1: The function set of GPPNM

1.	$f_1(x)=a_0+a_1x_1+a_2x_2+a_3x_1x_2$
2.	$f_2(x)=a_0+a_1x_1+a_2x_2$
3.	$f_3(x)=a_0+a_1x_1+a_2x_2+a_3x_1^2+a_4x_2^2$
4.	$f_4(x)=a_0+a_1x_1+a_2x_1x_2+a_3x_1^2$
5.	$f_5(x)=a_0+a_1x_1+a_2x_2^2$
6.	$f_6(x)=a_0+a_1x_1+a_2x_2+a_3x_1^2$
7.	$f_7(x)=a_0+a_1x_1+a_2x_2^2+a_3x_2^2$
8.	$f_8(x)=a_0+a_1x_1^2+a_2x_2^2$
9.	$f_9(x)=a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_1^2+a_5x_2^2$
10.	$f_{10}(x)=a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_1^2$
11.	$f_{11}(x)=a_0+a_1x_1+a_2x_1x_2+a_3x_1^2+a_4x_2^2$
12.	$f_{12}(x)=a_0+a_1x_1x_2+a_2x_1^2+a_3x_2^2$
13.	$f_{13}(x)=a_0+a_1x_1+a_2x_1x_2+a_3x_2^2$
14.	$f_{14}(x)=a_0+a_1x_1+a_2x_1x_2$

15.	$f_{15}(x)=a_0+a_1x_1x_2$
16.	$f_{16}(x)=a_0+a_1x_1x_2+a_2x_1^2$

Fig.1 shows the representation of individual in the population of GPPNM. For each function node in the tree, its coefficients are estimated by a rapid recurrent least squares (RLS) method to avoid the need to search for their values [13]. Its output is calculated according to the two arguments and mathematic function. Then it becomes one of the arguments of upper node. This process continues to the root node to generate final output F_{GP}.

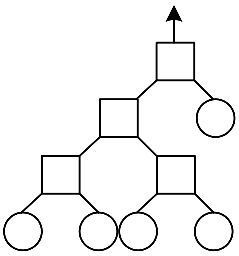

Fig. 1 Tree-like individual in the GPPNM

3.2 Fitness calculation

Predicted squared error (PSE) criterion is used in the GPPNM to evaluate each individual as

$$PSE = FSE + C*(2K/n)*\sigma^2 \quad (1)$$

Where FSE is the fitting squared error on the training data, C is a complexity penalty multiplier selected by the user, K is the number of model coefficients, n is the number of samples in the training set, and σ^2 is a prior estimation for the variance of the error obtained with the unknown model [9]. As the model becomes more complex relative to the size of the training set, the second term increases linearly while the first term decreases. PSE goes through a minimum at the optimum model size that strikes a balance between accuracy and simplicity. The user may optionally control this trade-off using the C parameter. Larger values than the default value of 1 lead to simpler models that are less accurate but may generalize well with previously unseen data, while lower values produce more complex networks that may overfit the training data and degrade actual performance.

3.3 Genetic operators

The reproduction, crossover and mutation operators are used to generate new individual of population. The reproduction operator directly keeps the individual to next generation without any change. The crossover operator exchanges parts of two individuals in the randomly selected crossover nodes while the new individual with small size is selected. The mutation operator varies the selection node with a new function node, terminal node or small size tree. All of the three operators are applied with predefined probabilities.

3.4 GPPNM-based Diagnosis System

Fig.2 displays the flowchart of GPPNM. The proposed GPPNM-based diagnosis system is constructed as follows.

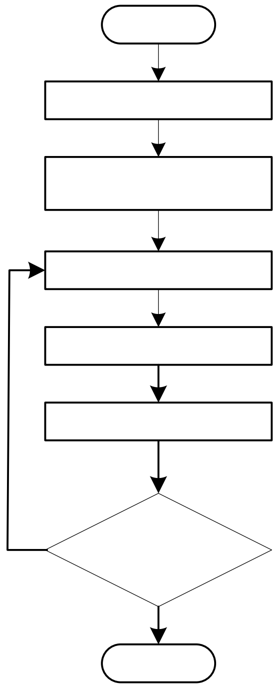

Fig. 2 Flowchart of GPPNM construction

3.4.1 Step 1 (DGA data collection)

Collect the historical insulation fault records with characteristic gases concentration, fault types and the recommendatory maintenance advices.

3.4.2 Step 2 (Initialization of GPPNM)

This step includes the selection of function set, terminal set, tree-generation algorithm, initialization of population, probabilities of reproduction, cross-over and mutation and so on.

3.4.3 Step 3 (Genetic operation)

Individuals in the old generation undergo genetic operations to generate new ones according to predefined probabilities.

3.4.4 Step 4 (Fitness calculation)

F_{GP}

In this step, the coefficient of each function node is first estimated by RLS method. Then fitness of individual is calculated by Equation 1.

3.4.5 Step 5 (Selection process)

The better the fitness of individual is, the more chance it survives in the new generation by a

f_2

f_{11}

tournament selection strategy. The new individual will replaced the one with poor fitness in the process of evolution.

3.4.6 Step 6 (Checking the stop condition)

While the best individual's fitness value satisfies stop criteria or max generation is arrived, the evolution process then stops and the best individual is output as solution to the diagnosis system; otherwise repeats step 3-6.

4 Hierarchical Classification Strategy For Insulation Fault Types

There are 4 insulation fault types in our diagnosis database. Due to the complicated structure and operation condition of power transformers, the reasons that cause insulation failures are complex. In order to decrease misclassification and provide diagnosis information step by step, a hierarchical classification strategy with four main fault types is proposed in Figure 2 where L.E.D., H.E.D., L.M.O., H.O. represent low energy discharge, high energy discharge, low or mediate temperature overheating, and high temperature overheating, respectively. The GPPNM1 firstly classifies the energy discharge faults and thermal faults. Then GPPNM2 classify the L.E.D. and H.E.D. while GPPNM3 the L.M.O. and H.O.

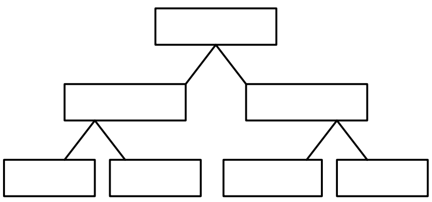

Fig. 2 A hierarchical classification strategy for insulation fault types

5 Numeric Test

The proposed GPPNM-based diagnosis system has been applied to DGA data in our database. Table 2 lists the composition of training set and verifying set. We use two kinds of feature variables of DGA data in our diagnosis system. Case 1 selects the three extensively used gas ratio C_2H_2/C_2H_4, CH_4/H_2, and C_2H_4/C_2H_6 [4]. Case 2 uses the relative percentage of above five gases' concentrations.

Table 3 lists the parameters adopted in the GPPNM. They are generally selected to keep the model from complexity and save computational time.

Table 2: Composition of the training set and verifying set

Fault case	Energy discharge		Thermal fault	
	L.E.D.	H.E.D.	L.M.O.	H.O.
Training set	34	60	42	90
Verifying set	23	41	28	60

Table 3: Parameters in the GPPNM

Parameters	Value
Population size	100
Max nodes	40
Max generation	100
Function set	List in Tab. 1
Terminal set	Case 1, case 2
Probability of reproduction	0.1
Probability of crossover	0.7
Probability of mutation	0.2
Tournament size	7
complexity penalty multiplier	1

Table 4 compares the performance of GPPNM with other existing methods on the DGA data. The results show that the proposed GPPNM has a better average performance than the other methods in both cases. It also shows that the more the feature variables are, the better the performance is. The shortage of ANNs and SOPN methods is that structures of these models should be determined in advance while it mainly relies on prior knowledge of authors [6, 10]. Our GPPNM possess of global search ability to automatically find the better structure which ensures better performance.

Table 4: Performance of different methods

Case	Methods	Total accuracy (%)
Case 1	GPPNM	88.2
	SOPN	87.9
	ANNs	87.6
	IEC	68.2
Case 2	GPPNM	92.2
	SOPN	89.4
	ANNs	89.1
	IEC	68.2

6 Conclusion

This paper proposes a GPPNM-based diagnosis system to detect the insulation fault types of power transformers. The highly nonlinear relationship of DGA data and fault types is described by a tree-like polynomial networks with a set of second-order polynomials. The advantage of proposed method is that its ability to evolve accurate and predictive polynomial models. The application on insulation fault diagnosis has proved its validity.

References:

[1] Interpretation of the analysis of gases in transformers and other oil-filled electrical equipment in service, *IEC Publication 599*, First Edition, 1978.

[2] E. Dornenburg and W. Strittmater, Monitoring oil cooling transformers by gas analysis, *Brown Boveri Review*, Vol. 61, No. 5, 1974, pp. 238-274.

[3] R.R. Rogers, IEEE and IEC code to interpret incipient faults in transformers using gas in oil analysis, *IEEE Trans. on Electrical Insulation*, Vol. 13, No. 5, 1978, pp. 348-354.

[4] ANSI/IEEE Std C57.104-1991, IEEE guide for the interpretation of gases generation in oil-immersed transformers, *IEEE Power Engineering Society*, 1992.

[5] H.T. Yang and C.C. Liao, Adaptive fuzzy diagnosis system for dissolved gas analysis of power transformers, *IEEE Trans. on Power Delivery*, Vol.14, No. 4, 1999, pp. 1342-1350.

[6] Y. Zhang, X. Ding, Y. Liu, P.J. Griffin, An artificial neural network approach to transformer fault diagnosis, *IEEE Trans. on Power Delivery*, Vol.11, No. 4, 1996, pp.1836-1841.

[7] Z. Wang, Y. Liu and P.J. Griffin, A combined ANN and expert system tool for transformer fault diagnosis, *IEEE Trans. on Power Delivery*, Vol. 13, No. 4, 1998, pp. 1224-1229.

[8] A.G. Ivakhnenko, Polynomial theory of complex system, *IEEE Trans. System Man Cybernetics*, Vol. SMC-1, No.4, 1971, pp. 364-378.

[9] S.J. Farlow, *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, Marcel-Dekker, New York, 1984, pp.87-103.

[10] Hong-Tzer Yang and Yann-Chang Huang, Intelligent Decision Support for Diagnosis of Incipient Transformer Faults Using Self-Organizing Polynomial Networks, *IEEE Trans. on Power Delivery*, Vol. 13, No. 3, 1998, pp. 946-952.

[11] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Cambridge, MA: MIT Press, 1992.

[12] Nikolay Y. Nikolaev and Hitoshi Iba, Regularization Approach to Inductive Genetic Programming, *IEEE Trans. on Evolutionary Computation*, Vol. 5, No. 4, 2001, pp. 359-375.

[13] D.K. Fadeev and V.N. Fadeeva, *Computational methods of Linear Algebra*, W.H. Freeman: San Francisco, CA, 1963.