Advanced Materials Research
ISSN: 1662-8985, Vols. 354-355, pp 1022-1026
doi:10.4028/www.scientific.net/AMR.354-355.1022

Online: 2011-10-07

Transformer Fault Diagnosis Based on Gene Expression Programming Classifier

Zhuo Dong[1, a], Yongli Zhu[1, b]

[1] North China Electric Power University, Baoding, Hebei, 071003, China

[a]dz1223@126.com, [b]yonglipw@heinfo.net

Keywords: gene expression programming, DGA, data normalization, transformer fault diagnosis

Abstract. Gene Expression Programming (GEP), which is suitable for transformer fault diagnosis Classification, is combined with transformer oil dissolved gas analysis (DGA), and also a method of transformer fault diagnosis based on self-adaptive GEP classification algorithm is proposed. We choose 400 groups of DGA measured data which includes a variety of failure and does not redundant as the training samples and test samples of the GEP classifier. A large number of diagnostic examples show that the proposed self-adaptive classification GEP is suitable for transformer fault diagnosis, and its performance is better than using Naive Bayes (NB) classifier, BP network and Immune classification.

Introduction

At present, the technology of dissolved gas analysis in transformer oil (referred to as DGA) is an important monitoring measure used for oil-filled electrical equipment in the power system. The DGA technology can detect early failures that exist in the transformer. In China, over 50% of the transformer faults are detected by the result of DGA test according to statistics. Usually there are many gases parameters affecting transformer faults, but it is impossible for us to build an equation system including all relevant parameters in the fault diagnosis model, and the high correlation between the various gases parameters will have adverse effects on the accuracy and rate of fault diagnosis. In order to simplify the problem, and improve the diagnostic rate and accuracy, a kind of gene expression programming algorithm (GEP) based on data normalization is proposed and used for transformer fault diagnosis.

GEP Algorithm

GEP(Gene Expression Programming) is invented by C.Ferreira, a Portuguese scientist. It's a genetic algorithm based on genotype and phenotype [1, 2]. GEP is the inheritance and development of GA and GP, which combines the advantages of GA and GP, and has the stronger ability to solve problems, so that we can solve complex problems with simple codes. Due to those advantages, the speed of GEP is 100 ~ 60000 times higher than that of GA and GP [3].

The basic principle and structure of GEP. Firstly, the GEP algorithm generates a suitable search space randomly, each basic unit in this search space has a fitness value; Then, according to the fitness values, genetic operators are used to deal with the units and produce the next generation of search space; With the evolution going forward, we get the optimal solution to the given problem finally [4]。

The main genetic manipulation used in the evolution of GEP includes selection, crossover and mutation [5]. 1) Selection directly retains the better fitness individuals to the next generation refer to the selection probability Pc. This manipulation will homogenize the group. 2) Crossover is the major genetic operation to produce new individuals. Two individuals are selected from the search space according to the fitness value in this process, and two nodes in the individuals are selected randomly as the crossing points, and then the sub-trees is exchanged. 3) Mutation is to select a node of the individual based on mutation probability as the mutation node, and generates a new sub-tree randomly, and then uses the newly generated sub-tree to substitute the original one.

Property identification and data preprocessing

Determining the variables of attribute and the fault type. We selected H_2, CH_4, C_2H_6, C_2H_4 and C_2H_2 as the attribute variables[6, 7]. This model classifies transformer state into 5 types: Normal, Low Energy Discharge, High Energy Discharge, Low Temperature Superheated (<700℃), High Temperature Superheated (>700℃), where Low Energy Discharge stands for weak spark discharge [8, 9].

Establishment of training set and test set. Taking the transformer type, capacity, operating environment and other relevant factors into Consideration, we choose transformers with voltage level equal or below 220KV as research subjects.

We collect 500 groups of gas concentration data in transformer oil from large number of documents and the DGA records of transformer in corresponding on-site fault investigation. From which we select 400 groups as the fault samples which can reflect all kinds of faults without redundant. There are 250 groups of normal operation data, and 150 groups with apparent failure.

Data Normalization. Since the value of input gas data varies widely, GEP classifier may be insensitive to the smaller number of data if input original data directly, and important features will be difficult to obtain. In order to fully mining the effective information in original data, this paper adopts fuzzy technology for data normalization [10, 11]. Normalization can prevent excessive weight when properties with a larger initial value compared to properties with smaller initial value. After repeated experiment and comparison with other normalization methods, a new method of data normalization is presented in this paper. The expression of normalized value d_i is shown as (1):

$$d_i=1-1/\exp(2*(X_i/T_i)2) \quad (i=1，2，\cdots，5) \tag{1}$$

Xi is the original value of each kind of gas；

Ti is the attention value of the corresponding gas；

Model establishing and algorithm analyses

The training model for GEP classifier is shown in Figure 1.After the training, we take full advantage of GEP classifier in the classification accuracy, training time and adaptability, and other areas, and establish a fault diagnosis model based on multi-GEP classifier, which is shown in Figure 2.

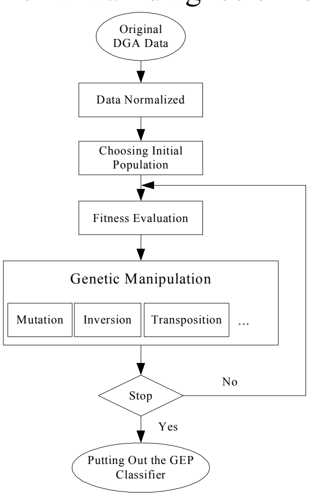

Fig. 1 The training model

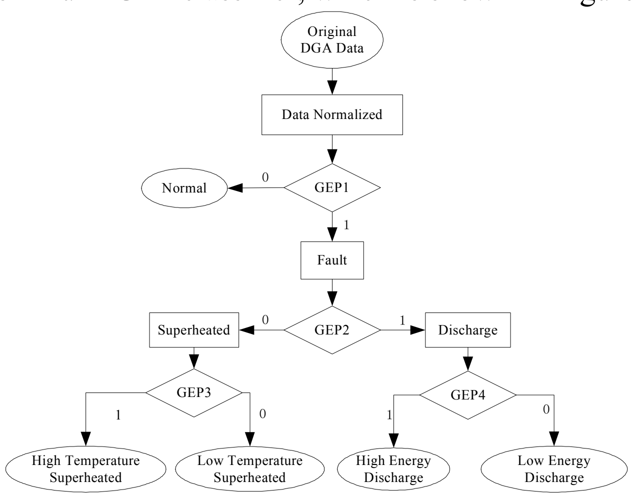

Fig. 2 Fault diagnosis model based on GEP classifier

Taking the evolution speed and training accuracy into account, and combining with the relevant information of the actual operation, the selected functions set F={+, -, *, /, sin, cos}.

GEP can reflect the relationship between training samples and the relevant factors, and adaptively identify factors associated with the load changes, then generate fault diagnosis model automatically. So we just select the Terminal set T={$d_1, d_2, d_3, d_4, d_5, C$}.Where d_1~ d_5 correspond to H_2, CH_4, C_2H_6, C_2H_4 and C_2H_2, and C corresponds to the constant set which generates randomly during the operation.

We assign the value of fitness threshold as 0.02, and set the critical point of GEP classifier as 0.5 (When the threshold value is greater than 0.5, GEP puts out 1, else puts out 0), and set the maximum evolution generation as Max Generation =1000, the two variables are both set as termination criterion. It means the program will stop running and the results will be taken out when one of the conditions is met. This method can ensure both the training accuracy and the training Speed.

The training condition of GEP classifier are shown in Table 1:

Table 1 Operating results of classifications

Classifier	Number of Training Samples	Training Accuracy
GEP1	200	95.50%
GEP2	107	96.97%
GEP3	54	100%
GEP4	53	100%

According to the logical relationship between classifications, and referencing to the process shown in figure 2, we connect the obtained GEP models with conditional statement "IF ... ELSE ... " and programming in the Visual Studio environment, then get the transformer fault diagnosis model based on multi-GEP classifier.

Diagnosis example

In order to validate the model accuracy in the actual diagnosis, we analyze the dissolved gas data gotten from the #2 main transformer of a thermal power plant. In this sample, the content of gases is shown in Table 2:

Table 2 The content of gases (μL/L)

H2	CH4	C2H6	C2H4	C2H2
12	290	147	335	0

Substitute the content of gases into Equation (3), normalize the data, then, we get the result shown in Table 3;

Table 3 The content of gases (μL/L)

d0	*d1*	*d2*	*d3*	*d4*
0.028389233	1	0.986724035	1	0

Refer to the process shown in Figure 2, we eventually get the result that the fault is Low Temperature Superheater, which is consistent with the actual situation.

The performance Comparison of diagnosis models

We test the 200 samples in the test sample set with GEP classifier which has been trained. Meanwhile, in order to verify the validity of the model, Bayesian classification, BP network and immune classification are also used in this paper to train samples on the same training samples, each obtained their own diagnostic models and test the sample in test sample set with the models above, the test results of the four methods are shown in table 4.

Table 4 Performance comparison of the two classifications

Property Values	The Number of Samples	GEP Classifier		NB classifier		BP network		Immune Classification	
		Correct Number	Accu-racy	Correct Number	Accuracy	Correct Number	Accuracy	Correct Number	Accu-racy
C1	87	86	98.85%	71	81.61%	72	82.76%	74	85.05%
C2	29	27	93.10%	22	75.86%	23	79.31%	23	79.31%
C3	30	29	96.67%	23	76.67%	26	86.67%	25	83.33%
C4	28	28	100.00 %	21	75.00%	23	82.14%	23	82.14%
C5	26	25	96.15%	19	73.08%	20	76.92%	21	80.77%

The test results indicates that, the average probability of accurate assessment of multi-GEP classifier is 96.5%, while it is 78% of NB classifier, 82% of BP network and 83% of immune classification. The model established in this article is better than the other three. This is mainly because the GEP algorithm itself has evolved the properties through the evolutionary, and constructs function adaptively, and approximates the nonlinear mapping relationship between the details and the transformer fault, which makes GEP classifier able to reflect in the evolutionary expression tree of the differences of transformer DGA sample information. Thus validate that the transformer fault diagnosis model based on multi-GEP classifier established in this paper is efficient and practical.

Conclusions

We construct a transformer fault diagnosis model based on GEP classifier and data normalization in this paper, which is suitable for fault classification diagnosis. Experiments show that the GEP algorithm has good ability to approximate arbitrary nonlinear mapping relationship, adaptive evolution and construct functional relationship which is close to the optimal solution, it helps to dig deep the related information between property values in transformer sample data, so that more accurate results can also be obtained in the limited training samples.

Under the same set of training samples and test samples, by comparing the test results of GEP classifier, NB classifier, BP network and immune classification, it is easy to know that GEP classifier is much better than the other three methods in adaptability, classification performance and probability of accurate assessment etc.

References

[1]. Ferreira C. "Gene expression programming: A new adaptive algorithm for solving problems," Complex Systems, vol. 13, pp. 87-129, 2001.

[2]. LI Qu, CAI Zhihua, ZHU Li, et al. "Application of Gene Expression Programming in Predicting the Amount of Gas Emitted from Coal Face," JOURNAL OF BASIC SCIENCE AND ENGINEERING, vol. 12, pp. 49-54, 2004.

[3]. Zhang Liechao, Cai Zhihua, Chen Ansheng. "A Survey of SGA, GP, GEP," Computer Information, vol. 22, p. 185, 2006.

[4]. FAN Xin-qiao, HUO Li-min, HUANG Li-hua, et al. "Short-term load forecasting based on Genetic Programming," Journal of North China Electric Power University, vol. 34, pp. 48-49, 2007.

[5]. Huo Limin, Yin Jinliang, Fan Yunfei, et al. "Short-term load forecasting of countryside distribution network based on improved gene expression programming," Transactions of the CSAE, vol. 25, pp. 193-194, 2009.

[6]. WU Zhong-li, YANG Jian, ZHU Yong-li, et al. "Power transformer fault diagnosis based on rough set theory and support vector machines," Power System Protection and Control, vol. 38, pp. 582-83, 2010.

[7]. Li Jian, Sun Caixin. "STUDY ON THE MODELS OF FAULT DIAGNOSIS ABOUT POWER TRANSFORMER BASED ON DISSOLVED GASES ANALYSIS," Chongqing: Chongqing University, 2002.

[8]. Wei Gen, Caixin Sun. "Study on fault diagnoses methods of transformer DGA with fuzzy model hierarchy classification" ICMEP, pp. 167-171, 2000.

[9]. ZHOU Zhi-Hua, YIN Xu-Ri, CHEN Zhao-Qian, et al. "APPLICATION OF NEURAL NETWORK TO THE DETECTION OF POWER TRANSFORMER RUNNING STATE". ACTA AUTOMATICA SINICA, vol. 28, pp. 301-302, 2002.

[10]. LIU Xiao-tong. "Study on Data Normalization in BP Neural Network,". MECHANICAL ENGINEERING & AUTOMATION, vol. 3, pp. 122-123, 2010.

[11]. LIANG Jia-zheng, XUE Zhi. "Standardized Processing Based on Network Data" Communications Technologies, vol. 7, pp. 46-47, 2010.

Progress in Power and Electrical Engineering
10.4028/www.scientific.net/AMR.354-355

Transformer Fault Diagnosis Based on Gene Expression Programming Classifier
10.4028/www.scientific.net/AMR.354-355.1022