Chapter 8 Load Tap Changer Fault Diagnosis

On-load tap changers (OLTCs) are one of the most problematic components of power transformers. Detecting faults in OLTCs is one of the key challenges facing the power equipment predictive maintenance community. This chapter will address the issue with an artificial intelligence approach, where logistic analysis is used to find the principal gases related to the fault conditions and neural networks are used to improve the performance of the diagnosis. The developed techniques could be integrated into power transformer fault diagnosis systems.

8.1 Why treat load tap changer separately

Detecting problems in on-load tap changers (OLTCs) is one of the major concerns in recent years [Young93, Young94, Desk95, Pomi98, Anon98, Lsec98, Kram96]. Methodologies include temperature monitoring, infrared thermal image tests, contact resistance measurement, and dissolved gas-in-oil analysis (DGA). DGA is by far the most popular technique for on-line OLTC diagnosis. Typical diagnostic gases are the same for transformer incipient fault diagnosis, including hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO) and carbon dioxide (CO_2). Some researchers also tried other special gases but the results were not conclusive [Kram96].

Interpretation of gases in OLTCs is different than that from the main tank of power transformers in that there are often byproduct gases present under normal OLTC operations because of discharges and heating. So the goal is to discern between normal gassing behavior and abnormal behavior, such as that from excessive contact wear and damage to the diverter switch. This concerns a special problem of OLTCs named "coking", which comes essentially from carbonized oil. Excessive "coking" can cause a thermal "runaway" of the contact and diverter switch materials as a result of increased contact resistance.

Some experience with DGA based "coking" diagnosis for OLTCs has already been accumulated and some diagnostic techniques have been devised [Young93, Young94, Desk95]. However, there is not a concensus on how to interpret DGA results from OLTCs. For instance, one researcher concludes that the amount of acetylene in a sample could be of no value in diagnosing a "coking" condition [Young94], while another argued that it is a good indicator.

This chapter will introduce the artificial intelligence (AI) method for detecting OLTC "coking" problems. The methods used include logistic regression analysis and neural networks. The conclusions could be the basis for the development of diagnostic algorithms for future OLTC monitors and power transformer incipient fault diagnostic systems.

8.2 The data sets for studying OLTC "coking" diagnosis

OLTC gas-in-oil concentrations are greatly related to whether or not the gas space is vented. Under normal conditions, free-breathing OLTCs with or without a desiccant jar have comparable gas contents, while completely sealed OLTCs have much higher gas contents [Young94]. This study considers only the free-breathing OLTCs, but the method developed applies to completely sealed OLTCs as well.

Three data sets were collected for the study [Young93, Young94, Desk95]. These include a training data set (OLTC_ TRN), a testing data set with no doubt in OLTC condition (OLTC_TST1), and a testing data set with OLTC condition in doubt (OLTC_TST2). The OLTC_TRN, OLTC_TST1 and OLTC_TST2 have 182, 16 and 30 data samples, respectively. Each sample includes the concentrations of the seven diagnostic gases (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, CO and CO_2) and the condition descriptions from expert diagnosis or internal inspections. The concentrations are given in parts per million (ppm), ranging from 0 to over 10000ppm. The conditions are described as "normal", "light coking", "moderate coking", and "serious coking". Examples in OLTC_TST2 have uncertain condition descriptions.

Of the 182 training data samples, 161 are considered to be from "normal" operating OLTCs and 21 are indicative of "coking" problems. Of the 21 samples indicating "coking", 4 samples are "light coking", 5 samples are "moderate coking", and 12 samples are "serious coking". Because the number of samples indicating "coking" is significantly less than that of the "normal" samples, simulated samples were generated from the "coking" samples through varying their concentration values by a random percent within ±10%. There are 156 simulated "light coking" samples, 155 simulated "moderate coking" samples, and 156 simulated "serious coking" samples. Therefore the actual training data set has 628 data samples.

It is believed that the simulated samples are valid representatives of the faulty OLTC cases because they do not deviate far away from the real cases.

8.3 Logistic regression based OLTC "coking" diagnosis

Logistic regression analysis is basically a model fitting technique. The goal is to find the best representation of the output as a LOGIT function of the inputs. A LOGIT function is defined as:

$$\text{LOGIT}(p) = \text{Log}\left[\frac{p}{1-p}\right] = W_0 + W^T X \tag{8-1}$$

Where p is the output, W_0 is a constant (the intercept), X is the $1 \times N$ input vector, W is the $N \times 1$ slope vector, and T denotes a vector transform.

W_0 and W can be estimated using a regression algorithm from a training data set, where p is a Boolean output, has a value of either 0 or 1, and represents whether or not an event occurred. After training, Equation (8-2) can be used to derive the probability of an event. When p is larger than 0.5, an event can be flagged as occurred.

$$p = \frac{\exp\left(W_0 + W^T X\right)}{1 + \exp\left(W_0 + W^T X\right)} \tag{8-2}$$

Apply OLTC_TRN to Equation (8-1) and find the estimation of W_0 and W, the following LOGIT functions were obtained.

$$\begin{aligned}\text{LOGIT}(p_1) = &-3.8981 + 0.00382H_2 - 0.0255C_2H_6 + 0.0105C_2H_4 \\ &- 0.00181C_2H_2\end{aligned} \tag{8-3}$$

$$\begin{aligned}\text{LOGIT}(p_2) = &-0.9904 + 0.000312H_2 - 0.00344CH_4 + 0.00582C_2H_6 \\ &+ 0.0137CO - 0.00055CO_2\end{aligned} \tag{8-4}$$

$$\begin{aligned}\text{LOGIT}(p_3) = &-4.4838 + 0.00124H_2 + 0.000964CH_4 - 0.00016C_2H_4 \\ &- 0.00118C_2H_2 + 0.0263CO - 0.00404CO_2\end{aligned} \tag{8-5}$$

Here p_1, p_2 and p_3 are used to discriminate between "normal / coking", "light coking / (moderate and severe coking)" and "moderate coking / severe coking", respectively.

Test results for the training data set are shown in Table 8-1, where the percent success rate of detecting "normal / coking" is based on all of the tested samples, while percent success rate for the two other pattern assignments are based only on the pool of "coking" data samples. For

example, the 84.2% success rate of "moderate coking / severe coking" diagnosis is based on 311 "(moderate and severe coking)" data samples.

Table 8-1 Testing success rates of logistic regression based OLTC "coking" diagnosis

Pattern assignment	Testing success rate (%)
Normal / coking	100
Light coking / (moderate and severe coking)	77.5
Moderate coking / severe coking	84.2

The 100% success rate of "normal / coking" is very impressive. The other two rates are also fairly high. The overall success rate is 89%.

Equation (8-3), (8-4) and (8-5) also identified the crucial gases for "coking" diagnosis. This will help resolving the disagreements on crucial diagnostic gases. These consent are normal among researchers but must be clarified for industrial applications. For instance, although most people recognized the relevance of C_2H_6 and C_2H_4 to "coking", there was argument on the importance of C_2H_2 [Young93, Young94, Anon98]. This must be settled.

Equation (8-3) clearly shows the importance of C_2H_2 in the detection of "coking". This is reasonable because C_2H_2 is a key gas indicator of both arcing and high temperature overheating conditions. Equation (8-4) implies that "light coking" corresponds to low-to-middle temperature overheating of OLTC oil. Equation (8-5) confirms that "severe coking' can be related to arcing. An interesting thing is the CO and CO_2 involvement in Equation (8-4) and (8-5). These gases may be from the decomposition of supporting insulation or the degradation of mineral oils in a high oxygen environment. The amount and relative portion of oxide gases formed is dependent on the type of incipient fault condition and materials.

8.4 MLP based OLTC "coking" diagnosis

The neural network based "coking" diagnosis studies involves data scaling, network topology optimization, and activation function selection. The considered neural networks will be introduced first, followed by definition of training performance evaluation parameters and the studies.

8.4.1 Neural Network Specifications

Four types of neural networks were considered. These are 7-1-1, 7-4-1 and 7-4-4 multi-layer perceptrons (MLPs), and a modular network topology consisting of four N-1-1 or N-4-1 MLPs. The activation function of the MLPs could be one of the followings.

Linear: $\quad f_1(x) = x$ $\hfill$ (8-6)

Sigmoid: $\quad f_2(x) = \dfrac{1}{1 + e^{-x}}$ $\hfill$ (8-7)

Odd sigmoid: $\quad f_3(x) = \dfrac{1 + e^{-x}}{1 + e^{-x}}$ $\hfill$ (8-8)

For the 7-1-1 and 7-4-1 MLPs, a "coking" condition is represented by an output $y \in [0,1]$. There were several scenarios to assign the value of y, which is named "output configuration" hereafter. In one scenario, ("normal", "light coking", "moderate coking", "severe coking") was represented by (0, .5, .75, 1). In another scenario, it was assigned (0, .7, .85, 1). These two scenarios correspond to a sigmoid activation function in the output layer. If an odd sigmoid activation function was used, (-1, 0, .5, 1) was assigned.

For the 7-4-4 MLP, a "coking" condition is represented by a 1×4 vector **y**, where **y** = [1 -1 -1 -1] represents "severe coking"; **y** = [-1 1 -1 -1] represents "moderate coking"; **y** = [-1 -1 1 -1] represents "light coking"; and **y** = [-1 -1 -1 1] represents "normal" condition.

The seven inputs for the above MLPs are necessary because we wanted to discern between different coking conditions using a single MLP.

Shown in Figure 8-1 is the topology of the modular network. Here the single complex task was assigned to three simple N-1-1 MLPs or three N-4-1 MLPs. The number N of inputs for each module MLP could be different based on its purpose. N is selected according to the results of the logistic regression analysis. A knowledge based output combination engine was used to combine the outputs of the three MLPs. Advantages of this topology include fast training, combination of human expertise, and flexibility of selection and separate training of the individual MLPs.
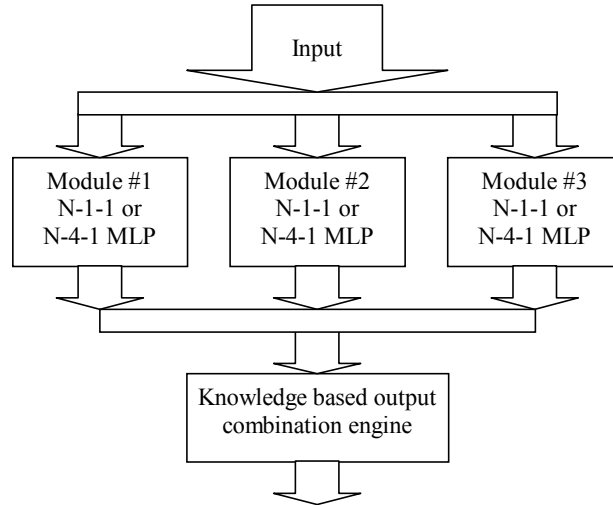
Figure 8-1 Modular network topology

8.4.2 Neural Network Training Evaluation Parameters

To compare the performance of different neural network topologies, two parameters are defined for the training evaluation. One is success rate (*SR*) and the other is informative index (*IFID*).

SR is a strictly defined parameter. For single output networks (7-1-1 MLPs and 7-4-1 MLPs) with a sigmoid activation function in the output layer, its value for each data sample is shown in Table 8-2, where *d* is the ideal output of the network and *y* is the real output. Overall *SR* for the training data set is the mean of all the *SR*s for each data sample.

Table 8-2 *SR* Definition for Single Output Neural Networks

d	$y \in$ [0, .25]	$y \in$ (.25,.625]	$y \in$ (.625,.875]	$y \in$ (.875,1]	d	$y \in$ (0,.35]	$y \in$ (.35,.775]	$y \in$ (.775,.925]	$y \in$ (.925,1]
0	1	0	0	0	0	1	0	0	0
0.5	0	1	0	0	0.7	0	1	0	0
0.75	0	0	1	0	0.85	0	0	1	0
1	0	0	0	1	1	0	0	0	1

IFID is a parameter to provide practical information. Because a "coking" condition is much different from a "normal" condition, and the differences between "coking" conditions are hard to distinguish, it is valuable to give credibility weights to the "coking" outputs. Therefore, for single output networks (7-1-1 MLPs and 7-4-1 MLPs) with a sigmoid activation function in the

output layer, the *IFID* for each data sample is defined in Table 8-3 where d is the ideal output of the network and y is the real output. The overall *IFID* for the training data set is the mean of all the *IFID*s for each data sample.

For a multiple output network (7-4-4 MLP), modular networks, 7-1-1 and 7-4-1 MLPs with an odd sigmoid activation function in the output layer, the definition of *SR* and *IFID* is similar to Table 8-2 and 8-3 but slightly different.

Table 8-3 Informative Index *IFID* Definition for Single Output Neural Networks

d	$y \in$ [0, .25]	$y \in$ (.25,.625]	$y \in$ (.625,.875]	$y \in$ (.875,1]	d	$y \in$ (0,.35]	$y \in$ (.35,.775]	$y \in$ (.775,.925]	$y \in$ (.925,1]
0	1	0	0	0	0	1	0	0	0
0.5	0	1	0.75	0.5	0.7	0	1	0.75	0.5
0.75	0	0.75	1	0.75	0.85	0	0.75	1	0.75
1	0	0.5	0.75	1	1	0	0.5	0.75	1

8.4.3 Data Scaling Studies

Because the neural networks may saturate from large inputs, the concentration values of the gases-in-oil must be scaled to small numbers. Assume X_{raw} is the gas concentration and u_i is the scaled input, three scaling schemes can be defined:

$$\begin{array}{llll} \text{a)} & u_i = X_{raw}/1000 & u_i \in [0, 1000) & (8\text{-}9) \end{array}$$

$$\begin{array}{llll} \text{b)} & u_i = X_{raw}/1000000 & u_i \in [0, 1) & (8\text{-}10) \end{array}$$

$$\begin{array}{llll} \text{c)} & u_i = \log_{10}(X_{raw}/1000+0.001)/3 & u_i \in [-1, 1) & (8\text{-}11) \end{array}$$

Table 8-4 gives the training performance comparison of the three schemes and the "no scaling" cases, where $\Phi_1(x)$ denotes the activation function of the middle layer. Activation function $\Phi_2(x) = f_2(x)$ was used for the output layer in all cases.

The table was obtained using a 7-1-1 MLP after 50 epochs of training. During each epoch of training, the data samples were provided to the training algorithm randomly. This ensures that the MLPs do not "forget" what they just learned.

Table 8-4 clearly shows that, in terms of *SR* and *IFID*, no data scaling and scaling with Scheme b) are totally unacceptable. When both of the activation functions are sigmoid, scaling with Scheme a) is almost the same as scaling with Scheme c). The best performance is obtained by scaling with Scheme a) and defining the activation functions to be $\Phi_1(x)= f_1(x)$, $\Phi_2(x)= f_2(x)$, i.e. linear in the middle layer and sigmoid in the output layer.

The schemes with good performance are shaded in Table 8-4 for clarity.

Table 8-4 **Comparison of Training Performance for Different Data Scaling Schemes**

Scaling Scheme	Output Configuration	$\Phi_1(x)$	SR	IFID
No scaling	(0, .5, .75, 1)	$f_1(x)$	0.2564	0.2564
	(0, .5, .75, 1)	$f_2(x)$	0.1242	0.5474
	(0, .7, .85, 1)	$f_1(x)$	0.3392	0.3702
	(0, .7, .85, 1)	$f_2(x)$	0.2484	0.5577
a)	(0, .5, .75, 1)	$f_1(x)$	0.7500	0.9064
	(0, .5, .75, 1)	$f_2(x)$	0.5653	0.8913
	(0, .7, .85, 1)	$f_1(x)$	0.7373	0.9033
	(0, .7, .85, 1)	$f_2(x)$	0.5653	0.8913
b)	(0, .5, .75, 1)	$f_1(x)$	0.2691	0.5681
	(0, .5, .75, 1)	$f_2(x)$	0.2484	0.5577
	(0, .7, .85, 1)	$f_1(x)$	0.2484	0.5629
	(0, .7, .85, 1)	$f_2(x)$	0.2484	0.5577
c)	(0, .5, .75, 1)	$f_1(x)$	0.2484	0.5577
	(0, .5, .75, 1)	$f_2(x)$	0.5573	0.8834
	(0, .7, .85, 1)	$f_1(x)$	0.2484	0.5577
	(0, .7, .85, 1)	$f_2(x)$	0.5621	0.8830

Actually, the performance of scaling with Scheme a) and defining the activation functions to be $\Phi_1(x)= f_1(x)$, $\Phi_2(x)= f_2(x)$ is so good that a single 7-1-1 MLP could be better than a modular network with four 7-1-1 MLPs, as shown in Table 8-5. This put the 7-1-1 linear-sigmoid MLP on the frontier position among examined neural network candidates.

Table 8-5 Comparison of Training Performance for a Single 7-1-1 MLP and a Modular Network with 7-1-1 MLPs

Network	Output Configuration	$\Phi_1(x)$	*SR*	*IFID*
7-1-1 MLP	(0, .5, .75, 1)	$f_1(x)$	0.7500	0.9064
7-1-1 MLP	(0, .7, .85, 1)	$f_1(x)$	0.7373	0.9033
Modular	$y \in [0,1]$	$f_1(x)$	0.7150	0.9021
Modular	$y \in [0,1]$	$f_2(x)$	0.7277	0.9050

8.4.4 Activation Function Studies

Besides the necessity of data scaling, Table 8-4 also shows the importance of the activation function selection. To further address this issue, Table 8-6 gives the training performance of different activation function assignment scenarios using a 7-4-1 MLP.

From Table 8-6 it is clear that an odd sigmoid activation function in the output layer is no good for a 7-4-1 MLP, while a linear activation function in the middle layer yields better results than a nonlinear one. This is consistent with what we observed in the data scaling studies, which basically says the linear/sigmoid function is suitable for the middle/output layer activation.

Table 8-6 Comparison of Training Performance for Different Activation Function Combinations

$\Phi_1(x)$	$\Phi_2(x)$	Output configuration	*SR*	*IFID*
$f_1(x)$	$f_2(x)$	(0, .5, .7, 1)	0.7611	0.9092
$f_1(x)$	$f_3(x)$	(-1, 0, .5, 1)	0.2484	0.5577
$f_2(x)$	$f_2(x)$	(0, .5, .7, 1)	0.5653	0.8913
$f_2(x)$	$f_3(x)$	(-1, 0, .5, 1)	0.2484	0.5577
$f_3(x)$	$f_2(x)$	(0, .5, .7, 1)	0.5885	0.8969
$f_3(x)$	$f_3(x)$	(-1, 0, .5, 1)	0.2484	0.5577

8.4.5 Modular Network Studies

The overall modular network topology is shown in Figure 8-1, in which Module #1 is for "normal / coking", Module #2 for "light coking / (moderate and severe coking)", and Module #3 for "moderate coking / severe coking". Table 8-7 shows the training performance of different

scenarios, where all the MLPs have a linear/sigmoid activation function in the middle/output layer.

In Table 8-7, Scenario #1 and #2 used all the gas-in-oil information to train the MLPs, Scenario #3 and #4 used some of the gas-in-oil concentrations (the selections are based on the results of logistic regression analysis). A slight increase can be seen in the success rates from Scenario #1 to #4, which means that simple MLPs tend to yield better performance, and knowledge based input MLP selection can improve the diagnostic performance.

Table 8-7 Comparison of Training Performance for Different Modular Networks

Scenario #	Module topology (MLP)			Overall
	#1	#2	#3	*SR* (%)
1	7-4-1	7-4-1	7-4-1	87.3
2	7-1-1	7-1-1	7-1-1	90.0
3	4-4-1	5-4-1	6-4-1	90.7
4	4-1-1	5-1-1	6-1-1	91.9

8.4.6 Discussions

Data scaling methods and activation functions greatly affect the performance of the diagnostic methods. If not properly selected, the AI technology could be less effective than a human expert. If properly determined, the performance of AI based diagnostic methods could be equal to or better than that of an expert. A systematic study is conducted to find the proper data scaling method. The result is clear and consistent with the AI based transformer oil DGA techniques presented in the previous chapters.

It is reasonable that a simple 7-1-1 MLP turns out to have similar performance to a modular network with several MLPs. This is because the results of logistic regression analysis (Table 8-1) are very close to that of modular networks (Table 8-7), and the equations of logistic regression are similar to a 7-1-1 MLP with a linear activation function in the middle layer.

Logistic regression analysis is the obvious choice if we are to write some heuristic equations for OLTC diagnosis. It also provides the guidance of selecting the inputs of neural networks for better performance (Table 8-7).

Modular networks are preferable if further performance improvement is needed. The improvement in this study is noticeable even if it is not very significant.

8.5 Apply the techniques to the testing data sets

A modular network based on 4-1-1 MLPs (Scenario #4, Module Topology #1 of Table 8-7) and Equation 8-3 were applied to the testing data set OLTC_TST1 and OLTC_TST2, the results are shown in Figure 8-2, where system output 1 means "coking".
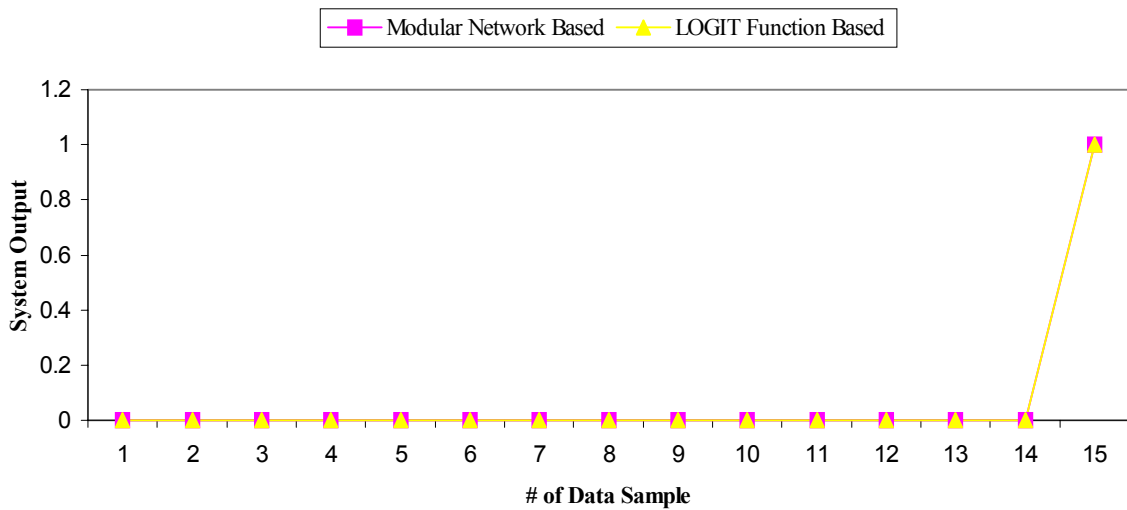
(a) Testing Results for Data Set OLTC_TST1
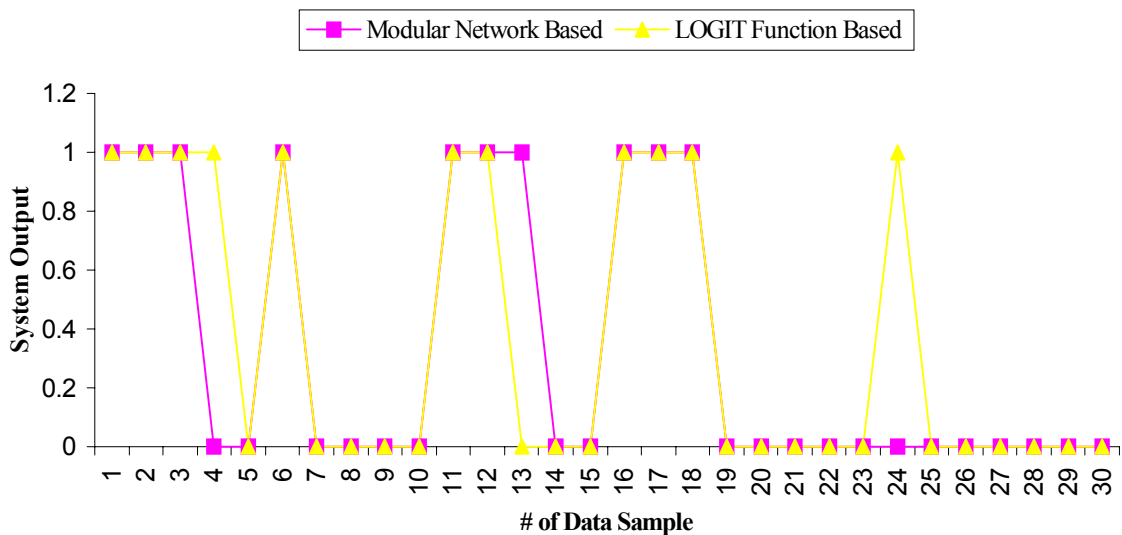
(b) Testing Results for Data Set OLTC_TST2

Figure 8-2 AI Based OLTC "Coking" Diagnosis

By comparing Appendix 9 data set OLTC_TST1 and Figure 8-1 (a), we see that both the MLP and the LOGIT function have great success in the diagnosis. They missed only one data sample (#14) in the 15 samples. If we examine the #14 data sample carefully, we can see that the gas-in-oil concentrations are actually very low and it is very hard to say that it is a "coking" case. Therefore the success rate of the two methods is almost 100%.

For data set OLTC_TST2, no success rate can be estimated because the uncertainty of the actual OLTC condition. But if we examine the two curves in Figure 8-2 (b), we can see they are very close to each other. Based on the confidence we have up to this point, if the two methods agree with each other, we follow them without question. If they don't, we need to consider the "light coking / (moderate and severe coking)"diagnosis also to reach the conclusion. This should be done using the true/false table in Table 8-8.

Table 8-8 "Coking" Diagnosis When MLP and LOGIT Function Fail to Agree

Conclusion	MLP Based		LOGIT Function Based	
	"Normal / Coking"	"Light Coking / (moderate and Severe Coking)"	"Normal / Coking"	"Light Coking / (moderate and Severe Coking)"
Normal	0	0	1	0
Coking	0	0	1	1
Coking	0	1	1	0
Coking	0	1	1	1
Normal	1	0	0	0
Coking	1	0	0	1
Coking	1	1	0	0
Coking	1	1	0	1

According to Table 8-8, the unmatched data samples #4, #13 and #24 in Figure 8-2 can all be diagnosed as "coking". This shows the effectiveness of modular networks, because the function of Table 8-8 is the responsibility of combination engine in Figure 8-1.

8.6 Summaries

Logistic regression analysis and neural networks were used to diagnose OLTC "coking" faults. Results show that the performance of logistic regression based diagnosis and neural network

based diagnosis are comparable. Logistic regression analysis can provide the guidance of selecting neural network inputs. Modular neural networks can improve the diagnostic performance. A linear/sigmoid activation function is preferred for the MLP neural network middle/output layer.