AN ARTIFICIAL NEURAL NETWORK
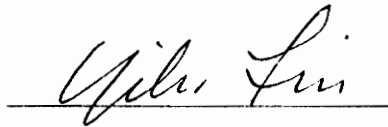# APPROACH TO TRANSFORMER FAULT DIAGNOSIS
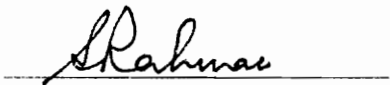
by

Yuwen Zhang

Thesis submitted to the Faculty of the

Bradley Department of Electrical Engineering of

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

IN

ELECTRICAL ENGINEERING

APPROVED:

_____

Dr. Yilu Liu, Chairman

_____   _____

Dr. Saifur. Rahman   Dr. Hugh F. VanLandingham

July, 1996

Blacksburg, Virginia

Keywords: Transformer, Artificial neural network, Dissolved gas analysis (DGA),

Fault detection and diagnosis.

An Artificial Neural Network

Approach to Transformer Fault Diagnosis

by

Yuwen Zhang

Chairman: Yilu Liu

Electrical Engineering

Abstract

This thesis presents an artificial neural network (ANN) approach to diagnose and detect faults in oil-filled power transformers based on dissolved gas-in-oil analysis. The goal of the research is to investigate the available transformer incipient fault diagnosis methods and then develop an ANN approach for this purpose. This ANN classifier should not only be able to detect the fault type, but also should be able to judge the cellulosic material breakdown. This classifier should also be able to accommodate more than one type of fault. This thesis describes a two-step ANN method that is used to detect faults with or without cellulose involved. Utilizing a feedforward artificial neural network, the classifier was trained with back-propagation, using training samples collected from different failed transformers. It is shown in the thesis that such a neural-net based approach can yield a high diagnosis accuracy. Several possible design alternatives and comparisons are also addressed in the thesis. The final system has been successfully tested, exhibiting a classification accuracy of 95% for major fault type and 90% for cellulose breakdown.

Acknowledgments

I would like to express my thanks to Dr. Yilu Liu, who as the chairman of the committee offered a great deal of advice during the research and the writing of the thesis. I would also like to thank my other committee members, Dr. Saifur Rahman and Dr. Hugh F. VanLangingham, for their guidance and time as members of my committee while I pursued my graduate degree. In addition, I would like to thank my fellow graduate students in the power area for their advice and friendship during my graduate studies. Finally, I want thank my family for the constant support and understanding that have made this all worthwhile.

TABLE OF CONTENT

List of Figures

List of Tables

List of Symbols

net_j = the net input to a node in layer j

o_j = the output of node j

θ_j = threshold or bias in sigmoidal activation function

θ_0 = shape modification factor in sigmoidal activation function

w_{ji} = weight linking node i to node j

i_{pi} = the i th training input feature of the pth training set

t_{pi} = the i th training output pattern corresponding to the pth training set

E_p = square error for training set p

E = average square error for P training sets

$\Delta_p w_{ji}$ = correction of w_{ji} due to pth training set

η = learning rate

α = momentum term

Chapter 1 Introduction

1.1 Transformer Fault and Protection

As a major apparatus in a power system, the power transformer is vital to system operation. Techniques for detection and diagnosis of incipient-fault are valuable. Failure diagnosis of power transformer in a power system in particular, is playing an important role in keeping the reliability of electricity supply. A transformer is subject to electrical and thermal stresses. These two stresses could break down the insulating materials and release gaseous decomposition products.

In order to minimize system outages, many devices have been evolved to monitor the service ability of power transformers. These devices, such as, Buchholz relays or differential relays, respond only to a serve power failure requiring immediate removal of the transformer from services, in which case, outages are inevitable. Thus, preventive techniques for early detection of faults to avoid outages would be valuable.

Under the effect of faults and/or failures of transformers (arcing, corona discharge, low energy sparking, serve overloading), the insulated oil and/or cellulose is decomposed and leads to the formation of gases that may be combustible or not, hence, an analysis of such gases make it possible to determine the occurrence of faults and/or failures associated with the dielectric materials involved, that is to determine the transformer's operating status.

Overheating, corona and arcing are three primary causes of fault related gases. Principally, the fault related gases commonly used are hydrogen (H_2), carbon monoxide (CO), carbon dioxide (CO_2), methane (CH_4), acetylene (C_2H_2), ethane (C_2H_6), and ethylene (C_2H_4).

1.2 Fault Detection with Dissolved Gases

Methods to identify fault conditions before they deteriorate to a severe state have attracted great research interest. While the most commonly applied methods is dissolved gas analysis (DGA), other useful methods include liquid chromatography, acoustic analysis, and transfer function techniques.

The analysis of dissolved gases is a powerful tool to diagnose developing faults in power transformers. Incipient electrical and thermal faults will release gases into the transformer oil. Thus, high concentrations of dissolved gas indicates possible fault conditions. The mechanisms of the gas generation are: (1)normal aging, (2)electrical discharges, or (3)excessive heating of the insulation structure due to eddy or load currents. The eddy current heating in item (3) is often caused by stray magnetic flux linking the metallic parts of the clamping structure, the core laminations, or the tank surface. Partial discharges in the transformer can usually be detected by electrical or acoustical methods, as well as the DGA method. However, the DGA analysis is often the only means of detecting stray flux heating deep within the transformer. Gas production is dependent on the type and energy of a fault. Many diagnostic criteria have been developed

for the interpretation of the dissolved gases. These methods would find the relationship between the gases and the fault conditions, some of which are obvious and some of which may not be apparent (hidden relationships). However, much of the diagnostics relies on experts to interpret the results correctly. It should be recognized that analysis of these gases and interpretation of their significance is , at this time, not mainly a science but an expert skill to variability. Even some diagnostic criteria are dependent on power transformer variables such as type, location, temperature, loading, gas levels, precursory failures and various other factors. It is possible that using different criteria will result in a different diagnosis result to same a transformer and operating condition.

In fact, there is not implicit relation between the various dissolved gas and the failure of transformer. In such case, it is worth to seek a new method to detect faults of transformer by means of dissolved gases information. This method does not need to use expert skill and to take into account some extraneous factor above. It only requires some gases levels.

New computer-aided techniques can consistently diagnose incipient-fault conditions for the novice and in some cases may provide further insight to the expert. Expert system and fuzzy-set approach have been developed to reveal some of the hidden relationships in transformer fault diagnosis[LIN93]. Expert system derives the decision rules from the previous experience. The analysis is improved by aggregating information from more than one of the available techniques.

1.3 Contributions of this Research

As mentioned above, the computer-aided expert system was developed to incorporate the available diagnostic criteria to improve the diagnosis performance, While the fuzzy-set represents the decision rules by using vague quantities. On the other hand, artificial neural network can learn and adapt to different statistical distributions, extract essential characteristics from input data, and do not require a physical model. Fault diagnosis is basically association of pattern of input data to one or more fault conditions. Thus, artificial neural network method (ANN)[DIN95][BHA93] is suitable for this task since the hidden relationships between the fault types and dissolved gases can be recognized by ANN through training process. The system described in this thesis utilize a neural network approach for the inspection of failed transformer. The major contributions of this research are as follows:

1. This thesis reviews and describes the relationship between the transformer fault and dissolved gases.

2. A two-step ANN classifier has been developed to detect the major fault type and cellulose degradation as well with high accuracy.

3. Different feature types and several different neural network topologies have been carefully assessed.

4. The performance of this ANN classifier was carefully verified.

5. The system accommodates different types of transformer since the normalized gas amount value(ppm value) is used.

6. A paper was presented at the IEEE/PES 1996 Winter Meeting, Baltimore, and to appear in IEEE Trans. Power Delivery, 1996[ZHA96].

Chapter 2 Literature Review

Attempts to diagnose the type of fault from the gases evolved from the oil/cellulose after the failure of mineral oil immersed power transformers started 50 years ago[ROG78], and had developed by 1956, into a detailed assessment of the fault from the gases collected in the Buchholz relay. It was quickly realized that if gases are evolved from the oil sufficient to operate a Buchholz relay, then slowly developing faults would also be producing decomposition gases which would be dissolved in the oil and only appear in the Buchholz at the end of a complicated system of interchange between the gases contained in bubble rising to the surface and the less soluble atmospheric gases dissolved in the oil. It should thus be possible by analyzing the gases dissolved in the oil or in a nitrogen cushion to detect former. Originally, indications of the type of fault occurring where developed from past history, or from experiment work on oil filled samples where the gases collected were differentiated as a major component of, or a percentage of the total gas accumulated.

From 1968, the C.E.G.B.(Central Electricity Generating Board)[ROG78] started regularly monitoring by chromatorgrapic analysis of gas in oil samples from increasing numbers on a routine basis, and by 1970 over one thousand units of 132, 275, and 400 KV voltage rating were checked at least annually. From statistical assessment of the results, attempts were made to set gas levels whereby 90% of all transformers being below these levels, the assumption could be made that transformers with gas levels above these

"Norms" were under suspicion of containing incipient faults likely to cause future trouble in service. However, study of the analyses shows that all transformers in service, including lightly loaded units known to be operating satisfactorily, evolved hydrogen and the other simple hydrocarbon gases, albeit in very small quantities, e.g. less than 0.1 ppm methane per month of service. Generally the total gas content depended on the time in service and the load conditions.

In 1970, Dornenburg[FAL70] differentiated between faults of thermal or electrical origin by comparing pairs of gases with approximately equal solubilities and diffusion coefficients such as ethylene and acetylene where an increase in the ratio of ethylene/acetylene above unity indicated an electrical fault, and also used the ratio of methane/hydrogen to suggest an indication of an thermal fault if the ratio was greater than 0.1 or a corona discharge if the ratio was less than 0.1. This methods was recognized promising, in that it eliminated the effect of oil volume and it simplified the choice of unit since one is measure in one gas produced per unit of another.

A report produced in 1970 by Halstead but not published until 1973[HAL73], made a theoretical thermodynamic assessment of the formation of the simple decomposition gaseous hydrocarbons in mineral oil which suggested that on the basis of equilibrium pressures at various temperatures, the proportion of each gaseous hydrocarbon in comparison with each of the other hydrocarbon gases varied with the temperature of the point decomposition. This led to the assumption that the rate of evolution of any particular gaseous hydrocarbon temperature there would be maximum

rate of evolution of that gas and that each gas would attains maximum rate at a different temperature. This consideration led to the choice of four ratios for fault diagnosis.

The various interpretation schemes employed in Europe were summarized in a CIGRE Study Paper in 1975[CIG75], and a trilinear scheme was described by Duval in 1974[DUL74]. In order to establish the identification of actual faults, the CIGRE WG 15-01 assessed one hundred sets of analyses from transformers with known faults, contributed from the records of members of the working group. The results of considerable laboratory work on small specimens were provided by several members in order to assess the probable temperatures at which the ratios indicated significant change. In the light of these results and further theoretical assessments, the significant changeover values of the ratios for both electrical and thermal faults were then modified.

Because the ratio ethane/methane only indicated a limited temperature range of decomposition, but did not assist in further identifying the fault, it was deleted. The modified tables were included in IEC Document 10A(secretariat)53 and detailed in the 1977 Doble Conference [AND76].

Although the ratio method was dominated in DGA analysis, in these years, the fuzzy-set approach and expert systems have been used to incorporate various rules[LIN93] [TOM93].

The effect of cellulose deterioration in assisting in the diagnosis of the phenomena should not be ignored. The relationship between carbon monoxide and carbon dioxide

under normal or abnormal conditions depends on the amount of oxygen available to assist the thermal decomposition of cellulose.

Partial discharges of high energy density are generally accompanied by tracking of solid insulation and the production of measurable quantities of CO and CO_2. Similarly for insulated conductor overheating, the increase of the unsaturated hydrocarbons such as ethylene compared with the saturated ethane should also be accompanied by increased CO and CO_2. There is now a confirmatory test for furanic compounds which are unique to the degradation of cellulosic materials[GRI95].

Chapter 3 Transformer Faults and Dissolved Gases

3.1 Transformer Faults and Dissolved Gases

Like many diagnostic problems, diagnosis of an oil-immersed power transformer is a skilled task. A transformer may function well externally with monitors, while some incipient deterioration may occur internally to cause a fatal problem in the latter development. According Japanese experience[LIN93], nearly 80% of all faults are resulted from incipient deteriorations. Therefore, faults should be identified and avoided at the earliest possible stage by some predictive maintenance techniques.

Fault gases in transformers are generally produced by oil degradation and other insulating materials, e.g., cellulose and paper. Theoretically, individual dissolved gas concentration, gassing rate, total combustible gas (TCG) and cellulose degradation are all significantly increased. Different patterns of gases are generated due to different intensities of energy dissipated by various faults. The insulating materials within the oil-filled transformer could be decomposed into gases as the result of the stresses that are described before. Principal decomposed gases include H_2, CH_4, CO, CO_2, C_2H_2 , C_2H_4 and C_2H_6. Totally or partially dissolved into the oil, the gases present in an oil sample make it possible to determine the nature of fault by the gas types and their amount. By using gas chromatography to analyze the gas dissolved in a transformer's insulating oil, it becomes feasible to judge the incipient fault types. Hence, an analysis of such gases makes it possible to determine the occurrence of faults at an early stage.

3.2 Fault Types and Dissolved Gas Analysis (DGA)

Dissolved gas analysis (DGA) method is a widely used technique for transformer fault diagnosis. Different patterns of gases are produced because of the intensity of energy that is dissipated by the fault. Many interpretative methods based on DGA to diagnose the nature of incipient deterioration have been reported. Even under normal transformer operational conditions, some of these gases may be formed inside. The theory behind the methods was derived mainly from the Halstead's discovery. Halstead made a theoretical-thermodynamic assessment of the formation of the simple decomposition gaseous hydrocarbons[HAL73]. It's stated that the rate of evolution of any particular gaseous hydrocarbon varies with temperature, and that at a particular temperature there would be a maximum rate of evolution of that gas. Study of the Halstead thermodynamic equilibrium suggested that with increasing temperature, the hydrocarbon gas with maximum rate of evolution would in turn be methane, ethane, ethylene, and acetylene. A relationship between the gas generated by thermal decomposition and the heating temperature of insulating oil is shown in **Figure 3.1**.

Gas Generated (ppm) vs Heating Temperature(°C)
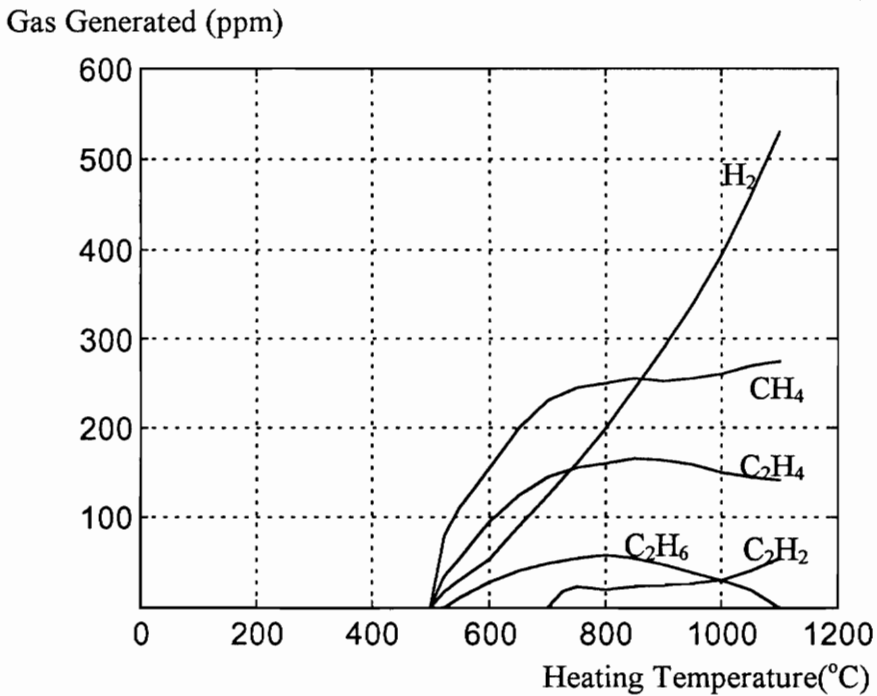
Figure 3.1 Relationship between gas generated and heating temperature

Halstead's work proved the existence of the relationship between fault temperature and the composition of dissolved gases. Tests were also conducted by others to examine how the amount of decomposed gases would vary with the temperature of insulating oil. It is shown H_2 rise steeply with temperature. Acetylene gas is the highest in generation at initiating temperature and its amount shows a steady rise with temperature. CH_4, C_2H_4 and C_2H_6 show a saturation or reduction with temperature. The conclusions obtained by the others[INO90] are the same as Halstead thermodynamic equilibrium.

3.3 Cellulosic Material Degradation

Overheating of paper, pressboard, and other cellulosic materials can be distinguished from overheating of oil as they generate carbon monoxide and carbon dioxide as by-products. However, infrequently carbon oxides are also generated from overheating of oil and aging. Knowledge of the gases that are generated as a "normal" transformer ages is necessary in order to determine when a transformer is behaving abnormally. Norms, generation rates, and the ratio of CO/CO_2 can be an indicator of cellulose degradation. The "normal" range of $0.07 < CO/CO_2 < 0.30$ was determined from measurements on ostensibly normally operating transformers[GRI95]. Rogers cites a mean ratio for CO/CO_2 of 0.143 as determined from a sample of over 1000 generator step-up and transmission-tie transformers in England. There is now a confirmatory test for furanic compounds which are unique to the degradation of cellulosic materials. The carbon monoxide is typically seven times less than the carbon dioxide, although this ratio can vary from 1:1 to greater than 1:20. Usually ratios tending towards 1:1(some people use 1:3 as a demonstration between normal and abnormal) indicate higher exposure temperature. In failed transformer in which arcing involves the cellulosic material, the CO/CO_2 is typically on the order of 1:1. In transformers which have inadequate cooling or are highly loaded, it is often the case that the CO_2 increases much more rapidly than the CO and, therefore, results in ratio of 1:10 to greater than 1:20. The use of CO/CO_2 ratios suffers from the presence of "accidental" CO_2 from air and from variations in O_2 concentration present in the transformer incipient fault area (high O_2 favors formation of CO_2). Also, oil, if

sufficiently heated in the presence of excess oxygen, will also generate some CO and CO_2. As a result, the detection of cellulose breakdown only based on the CO/CO_2 ratio may not give accurate diagnostic result. The other five hydrocarbon gases are considered together with CO/CO_2 ratio to help filter out the "disturbance" involved in the CO and CO_2.

Chapter 4 Dissolved Gas Analysis

4.1 Diagnosis Criteria

Efforts have been made to create simplified diagnosis criteria such as the key gas method and the ratio method which in essence are based on these variations in gassing characteristics with temperature to which the materials are subjected. Various diagnostic criteria can be found in the literatures, of which the best know are those of Rogers, Dornenburg[FALL70], Duval[DUV74], Doble[GRI88], Laborelec[GRI88], IEC[ROG78] and Expert systems[LIN93][TOM93]. Among them, the Rogers, Dornenburg and IEC criteria employ relationships between gas contents. The Rogers and Dornenburg criteria use four of these relationships, while the and IEC criterion uses only threes. Once these relationships have been determined, and depending on the result obtained, their values are turned into codes whose combinations are associated with specific faults. Since the number of possible combinations of such codes by fours and threes, depending on the criterion being used, by far exceeds the number of diagnoses that are available through these methods, many analyses do, in practice, fail to produce a diagnosis. In such cases, it is necessary to resort to the relative composition pattern which checks the percentage of each gas among the total combustible gases(TCG).

The Duval criterion calculates the relative amount of each gas in relation to their sum total. It was developed for the identification of faults in equipment which has already developed some problems; therefore it always produces a fault diagnosis, even when the

equipment is faultless, and should be used only for identification of a fault that has been previously determined by other criteria.

The most complete criterion is that of Laborelec, which takes into account, in addition to the individual composition of certain gases (H_2, CO, and C_2H_2), the sum of all hydrocarbons and the relationship of CH_4 and H_2 contents. The Laborelec criterion is the one that most often produces a diagnosis in the analysis of the various types of equipment.

The Doble, or relative composition pattern criterion, is the one normally used when other criteria fail to produce a diagnosis. The major advantage of other criteria over that of Doble's is that, by dealing with code combinations, they can easily be handled by a microcomputer. Since it does not involve definite numbers of codes, the Doble criterion is unfit for computers procedures. Laboratories normally use one of the others, resorting to the Doble criterion only if no clear diagnosis is obtained through other methods.

Since each of the available methods has advantages and drawbacks, when any individual method indicates a problem, that evidence may be clarified by other methods. For example, if DGA analysis indicates discharge activity, it may be prudent to perform an acoustic test before attempting any disruptive inspection. Thus, fuzzy information approach is used to integrating different transformer diagnostic methods[TOM93]. Fuzzy relations are combined with a fault tree to provide the analysis possible. DGA was implemented with satisfying results. Simple prototypes for other diagnostic methods were also implemented.

In recent years, an expert system[LIN93] for dissolved gas analysis has also been developed to incorporate various rules based on past experience and other heuristic methods to transform the results of chemical analysis of oil, together with the transformer, into a diagnosis of transformer health. Shown in **Figure 4.1**, the diagnosis is a weighted conclusion drawn from a number of data pertinent to the equipment, the reliability of which increases with the amount of information available from previous tests and the degree of experience of the laboratory performing the analysis. It makes the diagnostic method to be not only a statistical computation, but also a expert skill. This is a main drawback when adopting various diagnosis criteria.
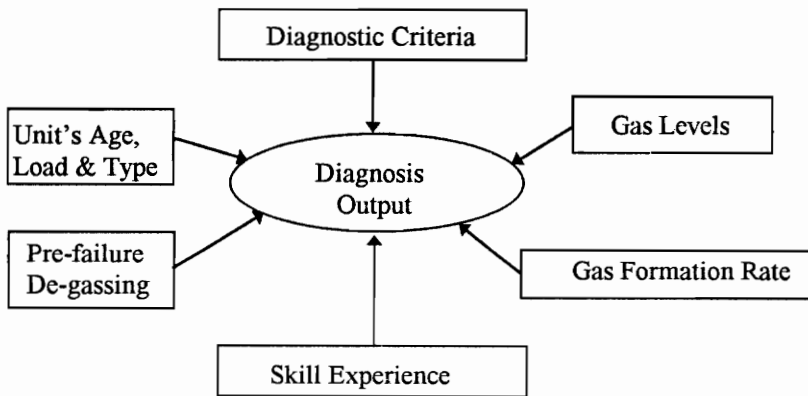
Figure 4.1 General diagnostic flow chart

4.2 Key Gas Method

Characteristic "Key gases" have been used to identify particular fault types[GRI88]. The characteristics of "Key gases" varies with the fault type, fault energy,

and insulating materials. The suggested relationship between key gases and fault types is summarized as below:

H_2----Corona

O_2 & N_2----Non-fault related gases

CO & CO_2----Cellulose insulation breakdown

CH_4 & C_2H_6----Low temperature oil breakdown

C_2H_4----High temperature oil breakdown

C_2H_2----Arcing

Nitrogen and oxygen are not typically considered diagnostic gas, but nevertheless are useful to assess the presence of leaks, overpressurization, or changes in pressure and temperature on the transformer system. Oxygen is one of the key promoters of oil and cellulose oxidation in transformer systems and, therefore, is generally maintained at low concentrations. Increases in oxygen suggest a leak in the system. A rapid decrease in oxygen with changes in oil properties and the generation of other gases may be indicative of overheating.

Changes in the nitrogen content can be due to changes in the temperature of the transformer in gas-blanketed. As the insulating oil heats and cools, there is a corresponding increase and decrease in the oil volume and relaxation of the gas space. The increase and decrease in pressure result in raising and lowering the dissolved nitrogen (and other gases) content in the oil. Excessive nitrogen can lead to bubble formation.

Excluding O_2 and N_2, there are seven fault related gases. The fault condition is indicated by the excessive generation of these gases. For example, acetylene is the key gas for arcing faults, hydrogen is the key gas for corona, and ethylene is the key gas for severe overheating. Excessive generation of these gases is indicative of fault conditions. By summarizing different combination, the fault types can be detected by using the following table.

Table 4.1. Key gas and fault types

Key Gases	Fault Types
H_2----High, $(CH_4+C_2H_6+C_2H_4)$---Not High, C_2H_2----Low	Corona or Arcing, No Cellulose Breakdown
C_2H_2----High, $(CH_4+C_2H_6+C_2H_4)$--- High, H_2----High	High Temperature Arcing
C_2H_4----High, $(CH_4+C_2H_6+C_2H_4)$--- High	Overheating or Local Overheating
C_2H_4, $(CH_4+C_2H_6+C_2H_4)$, H_2, CO, and CO_2---High	Overheating Involving Cellulose Breakdown
$(CH_4+C_2H_6+C_2H_4)$----High, H_2 and C_2H_2--- High, CO and CO_2---Normal	Overheating and Arcing Involving Cellulose Breakdown

There seems to be considerable differences over what is considered a normal transformer with acceptable concentrations of dissolved gases[ERD88]. For each type of condition, the level of concern at each total combustible gas level would vary. In the case

of electrical problems, erratic and rapid deterioration of the insulation is more likely and either high or very low levels of gases can be generated before failure occurs. In general, the overheating condition is slow and relatively high levels of combustible gases are generated as by-products. This permits ample time to plan and carry out remedial action.

For arcing and overheating conditions, the key gas method of identifying incipient faults can be refined. When arcing of low energy or of short duration occurs, there will be less overheating of the oil and, hence, less C_2H_4, CH_4 and C_2H_6 formed, even though the arc is of a sufficient hot-spot temperature to yield C_2H_2. Cooler transition zones about the arc will generate C_2H_4, CH_4 and C_2H_6 if sufficient time or energy is available to effect a large portion of the oil.

Overheating of oil can be further characterized based on thermodynamic considerations[HAL73]. As temperature increases, the concentration of gases increases as follows:

Low Temperature ($\cong 120\,°C$)	CH_4
	C_2H_6
	C_2H_4
High Temperature ($>700\,°C$)	C_2H_2

At a specific temperatures, one or two of these gases predominate while lesser quantities of the other gases are formed. In a practical situation, two scenarios are likely to occur which obscure this clear temperature discrimination. In one case, a thermal condition deteriorates over time, resulting in a changing spectrum of gases as the

temperature increases. In the other case, a large thermal gradient exists, giving rise to a wide spectrum of gases.

It should also be noted that other factors, such as high O_2, can result in additional gases being generated, for instance, CO and CO_2 which indicate the cellulose degradation.

Since this method does not give the numerical correlation, the diagnosis depends greatly on experience and therefore this technique is simple yet labor intensive.

4.3 Ratio Methods

Several analytical schemes have been developed which employ and compare ratios of characteristic gases generated under incipient fault conditions. The methods are used to determine the type of incipient fault condition similar to the key gas method. The advantages to the ratio methods are that they are quantitative, independent of transformer oil volume, and can be computer programmed.

Rogers, Dornenberg and IEC are the three most commonly used ratio methods. They employ the relationships between gas contents. The key gas ppm (parts per million) values are used in these methods to generate the ratios between them. The ppm value is defined below:

$$ppm = \frac{\left(gas \ in \ ft^3\right)(7.48)\left(10^6\right)}{gallons \ of \ oil} \qquad (4\text{-}1)$$

The ranges of the ratio are assigned to different codes which determine the fault types[ROG78]. Coding is based on experience and is always under modification. The gas ratio methods are promising because they eliminate the effect of oil volume and simplified the choice of units. Among these three methods, those of Dornenberg and Rogers are the most commonly used. In both methods, the same gases are used, however some of the ratios are different. Below, the gas concentration ratios used for each method are shown [ERD88].

Table 4.2 Gas ratios used for different criteria

ROGERS	DORNENBERG	IEC
CH_4/H_2	CH_4/H_2	C_2H_2/C_2H_4
C_2H_2/CH_4	C_2H_6/C_2H_2	CH_4/H_2
C_2H_4/C_2H_2	C_2H_2/CH_4	C_2H_4/C_2H_2
C_2H_2/C_2H_4	C_2H_2/C_2H_4	

Each gas concentration is taken in ppm. If the ration falls within a certain range, it is indicative of some fault conditions[ERD88]. **Table 4.3** displays the ratio method as proposed by IEC.

Table 4.3. Classification table by IEC

$\dfrac{C_2H_2}{C_2H_4}$	$\dfrac{CH_4}{H_2}$	$\dfrac{C_2H_4}{C_2H_6}$	Range of gas ratio (volume/volume)
0	1	0	smaller than 0.1
1	0	0	0.1-1
1	2	1	1-3
2	2	2	greater than 3

Characteristic Fault

0	0	0	Normal aging
*	1	0	Partial discharges of low energy density
1	1	0	Partial discharges of high energy density
1-2	0	1-2	Discharge of low energy (continuous sparking)
1	0	2	Discharge of low energy (with power follow through)
0	0	1	Thermal fault of low temp. less than 150 degree C
0	2	0	Thermal fault of low temp. between 150-300 degree C
0	2	1	Thermal fault of medium temp. between 300-700 degree C
0	2	2	Thermal fault of high temp. greater than 700 degree C

The codes for other ratio methods are available from the references. Ratio methods are limited in discerning problems when more than one type of fault occurs simultaneously. In addition, for some cases there is no diagnosis for a code as there are more possible combinations of the code than there are for the number of diagnosis. Ratio methods may used as an additional guideline to other diagnostic methods, such as key gas methods. Only as the amount of key gases increase to a certain value high enough to

indicated the existence of fault, the ratio methods are employed to diagnosis the fault types.

In ratio method, it is revealed that the fault type and severity are determined by the ratio values between different gases. Moreover, the overheating temperature can be estimated by the following equation:

$$T = 322 \lg\left(\frac{C_2H_4}{C_2H_6}\right) + 525^{\circ}C \qquad\qquad (4\text{-}2)$$

By estimating the overheating temperature, the severity of the fault can be determined.

4.4 Expert System and Fuzz-Set Approach

The fuzzy-set approach and expert systems have been used to incorporate various rules[LIN93][TOM93]. A knowledge base or a fuzzy membership function is selected based on the past experience. The fault diagnosis is a weighted conclusion drawn from a number of data pertinent to the equipment. Its reliability increases with the amount of information available from previous tests and the degree of experience of the laboratory performing the analysis. Therefore, the knowledge base required could be large and complex. The significant source in this knowledge base is the gas ratio method. Some limitations of this approach are overcome by incorporating the diagnostic procedure and the synthetic expertise methods. Furthermore, a data base adopted for certain gas records of transformers are incorporated into the expert system to increase the practical

performance. Uncertainty of diagnosis is managed by using fuzzy set concepts. Fuzzy set concept is used to handle uncertain norms thresholds, gas ratio boundaries and key gas analysis.

The knowledge base of this expert system is incorporated within the popular interpretative method of DGA, synthetic expertise and heuristic maintenance rules. Currently, these approaches are mainly depended on DGA methods since there are not much experience on non-DGA methods available. As more experience with non-DGA techniques is gained, the knowledge base will be refined. Current implementations of the non-DGA methods are trivial. While the performance is not highly sensitive to the various parameters in the fuzzy model, performance improvement can possibly be obtained by tuning the fuzzy parameters. Only modest improvements is expected due to the limited amount of fault data.

Chapter 5 Artificial Neural Network Approach

5.1 Artificial Neural Network

As discussed in the previous section, the diagnostic methods are not only computations, but also expert skills. By comparison, artificial neural network (ANN) method that simulates human nervous system may have new abilities such as learning, adaptation, self-organization that have big potential in abnormality diagnosis. Very complex systems can be characterized with very little explicit knowledge using ANNs. The relationship between gas composition and incipient-fault condition can be learned by the ANN from actual experience (through training samples). Obvious and not so obvious (hidden) relationships are detected by the ANN and used to develop its basis for interpretation of dissolved gas-in-oil data. Through training process, ANN can reveal complex mechanism that may be unknown to experts. In contrast, expert and fuzzy-set systems can only use explicit knowledge to establish knowledge base and fuzzy membership function selection. Theoretically, a neural network could represent any observable phenomenon.

ANN design includes selection of input, output, network topology (structure, or arrangement of nodes), and weighted connections of the nodes. Input feature (information) selection constitutes an essential first step. The feature space needs to be chosen very carefully to ensure that the input features will correctly reflect the

characteristics of the problem. The procedure is problem dependent, as described in detail in [DIN93].

Another major task of the ANN design is to choose network topology. This can be done experimentally through a repeated process to optimize the number of hidden layers (See **Figure 5.1**, a horizontal row of nodes is called a layer. The middle layers other than the inputs and outputs are called hidden layers) and nodes (computation units which take inputs from a lower layer, perform a specific mathematical function, and produce an output) according to training performance and prediction accuracy.

OUTPUT PATTERN

k   ⭘ ⋯⋯⋯ ⭘   OUTPUT LAYER

                         w(kj)

j   ⭘ ⭘ ⋯⋯⋯ ⭘   HIDDEN LAYER
                               (Internal representation units)

                         w(ji)

i   INPUT LAYER

INPUT PATTERN
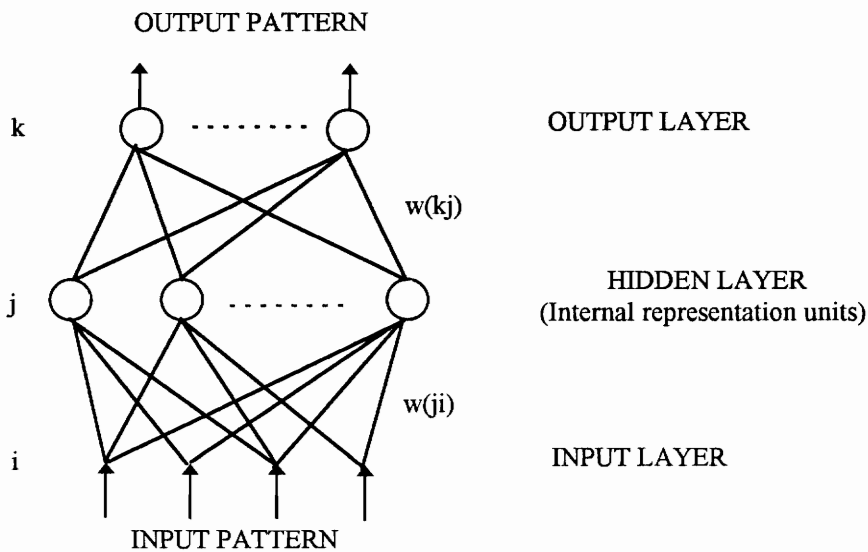
Figure 5.1 Schematic depiction of a semilinear feedforward connectionist net

Figure 5.2 illustrates an overall ANN design process with step by step adjustments to achieve desired structure and feature space. The corresponding connection weights are also determined in the process. In this thesis, back propagation learning

algorithm, alternatively known as the generalized delta rule, is used for ANN training[PAO89]. The back propagation learning algorithm consists of repeatedly passing the training set through the neural network until its weights minimize the output errors over the entire set.

Once the process is done, all weights will be retained and the ANN is now "trained" and ready for use. New samples will be fed into the trained ANN and predicted values can be readily obtained as shown in **Figure 5.2**.
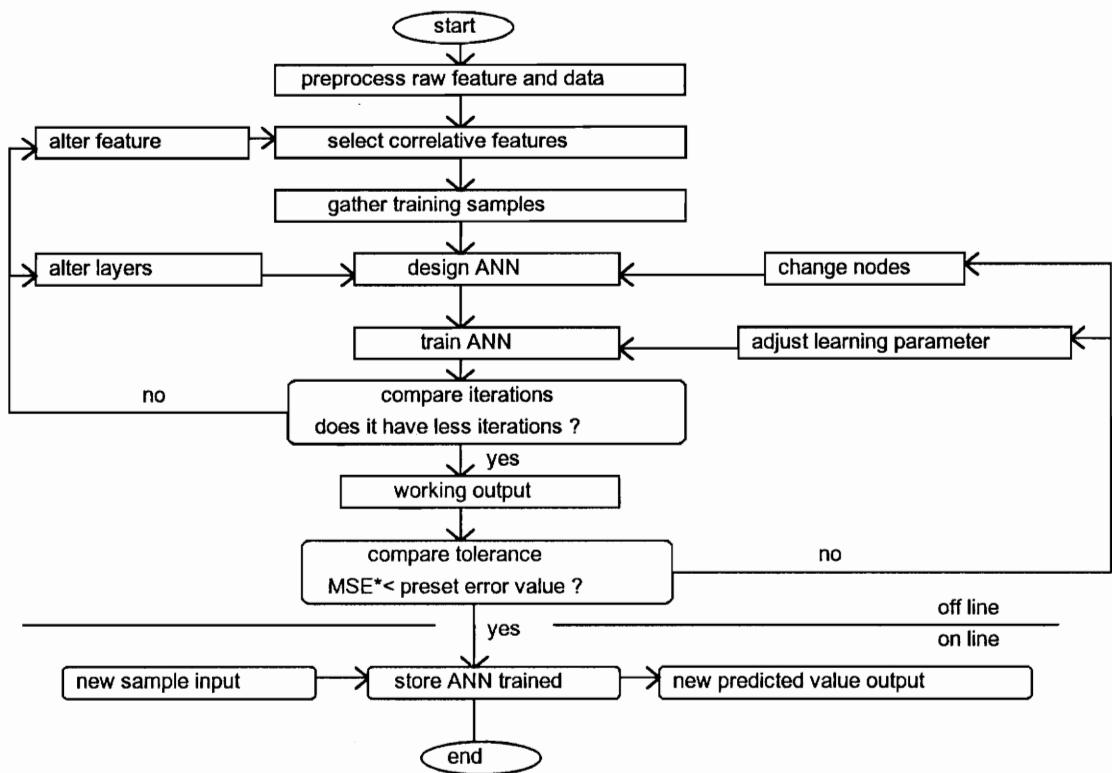
*MSE -------- Mean Square Error

Figure 5.2 ANN design and training flow chart

If a mapping function (relationship) exists between the input and output, finding weights and biases for the neural network is equivalent to finding the input/output relationship that describes the mapping function. Thus, neural network is especially appropriate and powerful when used to find relationships that are difficult to describe explicitly, since the weights and biases can be designed to represent a given function.

5.2 Back-Propagation Learning Algorithm

The back-propagation learning algorithm[PAO89] is an iterative gradient descent technique to minimize the mean square error between the actual output of an multilayer feedforward network and the desired output.

The semilinear feedforward net as reported by Rumelhart, Hinton, and Williams has been found to be an effective system for learning discriminates for patterns from a body of examples. The system architecture for such a net is illustrated schematically in **Figure 5.1**.

In general, such a net is made up of sets of nodes arranged in layers. The outputs of nodes in one layer are transmitted to nodes in another layer through links that amplify or attenuate or inhibit such outputs through weighting factors. Except for the input layer nodes, the net input to each node is the sum of the weighted outputs of the nodes in the

prior layer. Each node is activated in accordance with the input to the node, the activation function of the node, and the bias of the node.

Thus, in **Figure 5.1**, the components of an input pattern constitute the inputs to the nodes in layer i. The outputs of the nodes in that layer may be taken to be equal to the inputs, or we can take the opportunity to normalize those inputs in the sense that can be scaled to fall between the values of -1 and +1. The net input to a node in layer j is

$$net_j = \sum w_{ji} o_i \qquad (5\text{-}1)$$

The output of node j is

$$o_j = f(net_j) \qquad (5\text{-}2)$$

where f is the activation function.

For a sigmoidal activation function, we have

$$o_j = \frac{1}{1 + e^{-(net_j + \theta_j)/\theta_0}} \qquad (5\text{-}3)$$

In expression (5-3) the parameter θ_j serves as a threshold or bias. The effect of a positive θ_j is to shift the activation function to the left along the horizontal axis, and the effect of θ_0 is to modify the shape of the sigmoid. A low value of θ_0 tends to make the sigmoid take on the characteristics of a threshold-logic unit(TLU) which works as a digital switch, whereas a high value of θ_0 results in a more gently varying function. These effects are illustrated in **Figure 5.3**.
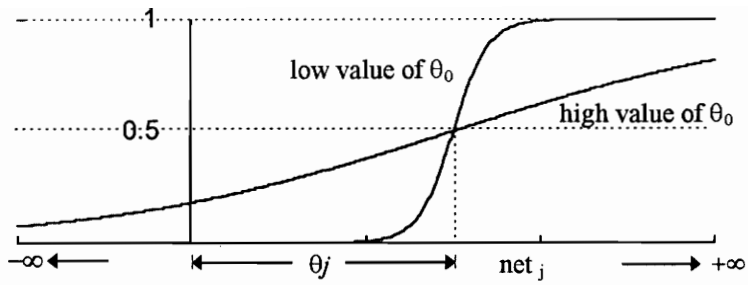
Figure 5.3 The sigmoidal activation function, with bias and shape modification

Continuing our description of the computational processes, we have for the nodes in layer k the input

$$net_k = \sum w_{kj} o_j \qquad (5\text{-}4)$$

and the corresponding outputs

$$o_k = f(net_k) \qquad (5\text{-}5)$$

In the learning phase of training such a net, we present the pattern $\underline{x_p} = \{i_{pi}\}$ as input and ask that the net adjust the set of weights in all the connecting links and also all the thresholds in the nodes such that the desired outputs t_{pk} are obtained at the output nodes. Once this adjustment has been accomplished by the net, we present another pair of $\underline{x_p}$ and $\{t_{pk}\}$, and ask that the net learn that association also. In fact, we ask that the net find a single set of weights and biases that will satisfy all the (input, output) pairs presented to it. This process can pose a very strenuous learning task and is not always readily accomplished.

31

In general, the outputs $\{o_{pk}\}$ will not be the same as the target or desired values $\{t_{pk}\}$. For each pattern, the summation of the square of the error is

$$E_p = \frac{1}{2}\sum_k (t_{pk} - o_{pk})^2 \tag{5-6}$$

and the average system error is

$$E = \frac{1}{2P}\sum_p \sum_k (t_{pk} - o_{pk})^2 \tag{5-7}$$

where P is the number of training sets and the factor of one-half is inserted for mathematical convenience at some later stage in our discussion.

In the generalized delta rule formulated by Rumelhart, Hinton, and Williams for learning the weights and biases, the procedure for learning the correct set of weight is to vary the weights in a manner calculated to reduce the error E_p as rapidly as possible. In general, different results are obtained depending on whether one carries out the gradient search in weight space on the basis of E_p or of E. In the former case, the corrections to the weights are made sequentially, on the basis of the learning to be carried out for the sequence of the patterns, one at a time. A true gradient search for minimum system error should be based on the minimization of expression (5-7). This latter procedure is feasible for adaptive pattern recognition, but is unlikely to correspond to processes carried out in biological neural nets.

In expression (5-6), we achieve convergence toward improved values for the weights and thresholds by taking incremental changes Δw_{kj} proportional to $-\partial E_p / \partial w_{kj}$, that is,

$$\Delta w_{kj} = -\eta \frac{\partial E_p}{\partial w_{kj}} \qquad (5\text{-}8)$$

However, E_p, the error, is expressed in terms of the outputs o_k, each of which is the nonlinear output of the nodes k. That is,

$$o_k = f(net_k) \qquad (5\text{-}9)$$

where net_k is the input to the kth node and by definition is the weighted linear sum of all the outputs from the previous layer:

$$net_k = \sum w_{kj} o_j \qquad (5\text{-}10)$$

The partial derivation $\partial E_p / \partial w_{kj}$ can be evaluated using the chain rule

$$\frac{\partial E_p}{\partial w_{kj}} = \frac{\partial E_p}{\partial net_k} \frac{\partial net_k}{\partial w_{kj}} \qquad (5\text{-}11)$$

Using expression (5-10), we obtain

$$\frac{\partial net_k}{\partial w_{kj}} = \frac{\partial}{\partial w_{kj}} \sum w_{kj} o_j = o_j \qquad (5\text{-}12)$$

We now define

$$\delta_k = -\frac{\partial E_p}{\partial net_k} \qquad (5\text{-}13)$$

and write

$$\Delta w_{kj} = \eta \delta_k o_j \qquad (5\text{-}14)$$

To compute $\delta_k = -\partial E_p/\partial net_k$, we use the chain rule to express the partial derivative in terms of two factors, one expressing the rate of change of error with respect to the output o_k, and the other expressing the rate of change of the output of the node k with respect to the input to that same node. That is, we have

$$\delta_k = -\frac{\partial E_p}{\partial net_k} = -\frac{\partial E_p}{\partial o_k}\frac{\partial o_k}{\partial net_k} \tag{5-15}$$

The two factors are obtained as follows:

$$\frac{\partial E_p}{\partial o_k} = -(t_k - o_k) \tag{5-16}$$

and

$$\frac{\partial o_k}{\partial net_k} = f'(net_k) \tag{5-17}$$

From which we obtain

$$\delta_k = (t_k - o_k)f'(net_k) \tag{5-18}$$

for any output-layer node k, we have

$$\Delta w_{kj} = \eta(t_k - o_k)f'(net_k)o_j = \eta\delta_k o_j \tag{5-19}$$

Circumstances are different if the weights do not affect output nodes directly. We still write

$$\Delta w_{ji} = -\eta\frac{\partial E_p}{\partial w_{ji}} = -\eta\frac{\partial E_p}{\partial net_j}\frac{\partial net_j}{\partial w_{ji}} = -\eta\frac{\partial E_p}{\partial net_j}o_i$$
$$= \eta\left(-\frac{\partial E_p}{\partial o_j}\frac{\partial o_j}{\partial net_j}\right)o_i = \eta\left(-\frac{\partial E_p}{\partial o_j}\right)f'_j(net_j)o_i = \eta\delta_j o_i \tag{5-20}$$

However, the factor $\partial E_p / \partial o_j$ can not be evaluated directly. Instead, we write it in terms of quantities that are known and other quantities that can be evaluated. Specifically, we write

$$-\frac{\partial E_p}{\partial o_j} = -\sum_k \frac{\partial E_p}{\partial net_k}\frac{\partial net_k}{\partial o_j} = \sum_k \left(-\frac{\partial E_p}{\partial net_k}\right)\frac{\partial}{\partial o_j}\sum_m w_{km}o_m$$
$$= \sum_k \left(-\frac{\partial E_p}{\partial net_k}\right)w_{kj} = \sum_k \delta_k w_{kj}$$
(5-21)

We see that, in this case,

$$\delta_j = f_j'(net_j)\sum_k \delta_k w_{kj}$$
(5-22)

This is, the deltas at an internal node can be evaluated in terms of the deltas at an upper layer. Thus, starting at the highest layer---the output layer---we can evaluate δ_k using expression (5-18), and we can then propagate the "errors" backward to lower layers. Summarizing, and using the additional subscript p to denote the pattern number, we have

$$\Delta_p w_{ji} = \eta \delta_{pj} o_{pi}$$
(5-23)

If the j nodes are output-layer nodes, we have

$$\delta_{pj} = (t_{pj} - o_{pj})f_j'(net_{pj})$$
(5-24)

However, if the j nodes are internal nodes, then we need to evaluate δ_{pj} in terms of δ_s at a higher layer; that is

$$\delta_{pj} = f_j'(net_{pj})\sum_k \delta_{pk} w_{kj}$$
(5-25)

35

In particular, if

$$o_j = \frac{1}{1 + \exp\left[\left(-\sum_i w_{ji} o_i + \theta_j\right)\right]} \tag{5-26}$$

then

$$\frac{\partial o_j}{\partial net_j} = o_j(1 - o_j) \tag{5-27}$$

and the deltas are given by the following two expressions:

$$\delta_{pk} = (t_{pk} - o_{pk}) o_{pk}(1 - o_{pk}) \tag{5-28}$$

$$\delta_{pj} = o_{pj}(1 - o_{pj}) \sum_k \delta_{pk} w_{kj} \tag{5-29}$$

for the output-layer and hidden-layer units, respectively.

The thresholds θ_j are learned in the same manner as are the other weights. We simply imagine that θ_j is the weight from a unit that always has an output value of unity. Also the derivative $\partial o_j / \partial net_j$, equal to $o_j(1 - o_j)$, reaches its maximum for $o_j = 0.5$ and, since $0 \leq o_j \leq 1$, approaches its minima as o_j approaches zero or one. Since the change in weight is proportional to this quantity, it is clear that weights that are connected to units in their midrange are changed the most. In some sense, these units are still uncommitted and are not certain whether to "turn on" or "turn off." The weights change rapidly under those conditions, and this feature probably contributes to the stability of learning procedure.

It is important to note that, for the activation function given by expression (5-26), a node can not have output values of 1 or 0 without infinitely large positive or negative

weights. Therefore, in learning mode, the values of 0.9 and 0.1 might suffice for specifying binary target output values.

In leaning w_{ji}, it is good practice to calculate $\Delta_p w_{ji}$ for each pattern in the training set of patterns, and to take

$$\Delta w_{ji} = \sum_p \Delta_p w_{ji} \qquad (5\text{-}30)$$

The leaning procedure therefore consists of the net starting off with a random set of weight values, choosing one of the training-set patterns, and, using that pattern as input, evaluating the output(s) in a feedforward manner. The errors at the output(s) generally will be quite large, which necessitates changes in weights. Using the back-propagation procedure, the net calculates $\Delta_p w_{ji}$ for all the w_{ji} in the net for that particular p. This procedure is repeated for all the patterns in the training set to yield the resulting Δw_{ji} for all the weights for that one presentation. The corrections to the weights are made and the output(s) are again evaluated in feedforward manner. Discrepancies between actual and target output values again result in evaluation of weight changes. After complete presentation of all patterns in the training set, a new set of weights is obtained and new outputs are again evaluated in a feedforward manner.

The net can be made to track the system error and also the errors for individual patterns. In a successful leaning exercise, the system error will decrease with the number of iterations, and the procedure will converge to a stable set of weights, which will exhibit only small fluctuations in values as further leaning is attempted.

There are several other issues we need to keep in mind when we implement such nets. There is, for example, the question of how the value of η is to be chosen. This is not a new or unusual problem; it is common to all steepest-descent methods of locating minima of function. As might be expected, a large η corresponds to rapid learning but might also result in oscillations. Rumelhart, Hilton, and Wiliams suggest that expressions (5-19) and (5-23) might be modified to include a sort of momentum term. That is we write

$$\Delta w_{ji}(n+1) = \eta(\delta_j o_i) + \alpha \Delta w_{ji}(n) \qquad (5\text{-}31)$$

Where the quantity $(n+1)$ is used to indicate the $(n+1)$th step, and α is a proportionality constant. The second term in expression (5-31) is used to specify that the change in w_{ji} at the $(n+1)$th step should be somewhat similar to the change undertaken at the nth step. In this way, some inertia is built in, and momentum in the rate of change is conserved to some degree. Examination of the system error e over a large number of steps in the iterative approach to solution will generally show that a finite α tends to dampen the oscillations but can also serve to slow the rate of learning.

If weight corrections are carried out after presentation of each pattern, the method is not truly a gradient search procedure. In addition, the value η needs to be small; otherwise, large excursions can take place in weight space.

That net must not be allowed to start off with a set of equal weights. It has been shown that it is not possible to proceed from such a configuration to one of unequal weights, even if the latter corresponds to small system error.

Another issue concerns the question of whether the system might get trapped in some local minimum or even at some stationary point, or perhaps oscillated between such points. Under such circumstances, the system error remains large regardless of how many iterations are carried out.

In a summary, at the beginning, all weights and biases are initialized to random values. The input values are presented, and the desired outputs are specified. The recursive algorithm, starting at the output nodes and working back to the hidden layer, adjusting weights until they converge and the error is reduced to an acceptable value. Back-propagation finds better solutions when momentum is used. Training time is shortened by the use of an adaptive learning rate [HOW92][ROG78].

5.3 Analysis of ANN Features and Topology

Since there are already various diagnostic criteria in use, the significance of using ANN is to achieve better diagnosis performance. It is also useful to distinguish the fault in cellulose from that in oil, which usually is a harder issue than the major fault type diagnosis.

To increase the accuracy of the ANN, it is necessary to choose the proper input features, output patterns and ANN topology. Overheating, corona and arcing are three major fault types. Since each of them could involve cellulose breakdown, there could be seven patterns to be identified, including the normal condition. For a given application, it

is usually not easy to decide what combination of those features has the optimal discriminatory power. This thesis compares the performance obtained using several feature types.

As mentioned in the previous chapters, five hydrocarbon gases and carbon oxide are fault related gases. Among them, CO and CO2 contribute much to the detection of cellulose breakdown. The input features must be selected from these seven gases. The ppm values of these gases are used to make the ANN accommodate transformers with different types.

Theoretically the above diagnosis and detection problem can be realized by one ANN using all 7 fault related gases as inputs. However, it is commonly accepted that the detection of cellulose breakdown is harder than major fault type diagnosis. It was found in a previous study that the ANN with CO_2 as one of the input feature has worse performance in major fault type (arcing, corona, overheating) diagnosis than that without CO_2[DIN95]. The reason is the amount of CO_2 is affected by the aging, transformer structure, environment and etc.. For example, the environment temperature has an unnegligible effect on the amount of CO_2, which may result 50% deviation in maximum. The value of CO_2 amount contains a lot of "disturbance" information which degrade the accuracy of overall fault diagnosis. However, CO and CO_2 are needed as the input features for the cellulose degradation detection since they are the primary gases formed from the degradation of cellulose.

For this reason, two ANNs are used to separate the cellulose condition detection from major fault type diagnosis. One ANN is trained only to classify the major fault types while another focuses on determining if the cellulose is involved.

The topology of a neural network has an impact on both the speed of convergence during the training, and the resulting classification accuracy. Because of this, a number of different network topologies were evaluated here. Let N be total number of total weights for a given network topology. Let n be the number of layers in the network, and L_i be the number of neurons in the ith layer of the network. The total number of weights can be expressed by:

$$N = \sum_{i=1}^{n-1} L_i L_{i+1} \tag{5-32}$$

The number of weight determines the complexity of the network. As different topologies are checked, it is reasonable to select the topologies with almost the same weight number, that is the same complexity.

5.4 Ten-fold Cross Validation

In order to compare the performance, we need to define the performance measurement first. The performance of the neural-net classifier can be measured by its classification accuracy. This is the ratio of the number of correctly identified cases to the total number of cases that were examined, The larger the set it examines, the more accurate the measurement can be. In a real classification problem, there is always only a

limited number of samples available. Therefore, a common practice is to split the samples into two groups, the training set and the testing set. When the training set is used to set the weight of the classifier, and the testing set is used to measure the performance of the classifier, it is called a single train-and-test experiment. When multiple train-and-test experiments are performed, a new classifier results from each training set. The estimated accuracy is chosen as the average of the accuracy for classifiers derived for the independently and randomly generated test partitions. Obviously, multiple train-and-test experiments can achieve a better measurement of the neural-net classifier than a single train-and-test experiment, by only using these limited samples[WEI91].

Given n samples, the most straightforward way of performing multiple train-and-test experiments is to train the classifier with $(n-1)$ samples and test it on the single remaining sample, and repeat this n times. This technique is called "leave one out." Because it is computationally extensive, a method called k-fold cross-validation is usually used to replace it. In our study, ten-fold cross-validation is used.

Ten-fold cross-validation is used to train and test the ANN, which is typically used in ANN accuracy verification. This is done as follows. Sample data are randomly divided into ten mutually exclusive test partitions of approximately equal size. Alternately, data of each set are tested by the ANN trained with the other nine set of data. This approach effectively increases the number of data for ANN testing. The obtained average accuracy reasonably represents the performance of the ANN classifier.

5.5 ANN for Major Fault Type Diagnosis

As discussed in the previous section, the six fault related gas H_2, CH_4, C_2H_2, C_2H_6, C_2H_4 and CO may be chosen as the input features, while the four output nodes can be four patterns: normal, overheating, corona and arcing (three major fault types). Since reducing the irrelevant input features and output nodes may affect the ANN performance, the ANN features and topologies will be carefully checked and refined in the present study to obtain the optimal design for the major fault diagnosis. The correlation between CO content and major fault types are checked since it is a cellulosic breakdown related gas and may not be tightly correlated with major fault types. It is also interesting to see if the "normal" output node is necessary, since no fault means normal condition. Details are given in this section.

In our study, five key gases H_2, H_4, C_2H_2, C_2H_6, C_2H_4 are chosen as input features. Both the cases with and without "CO" as input features have been studies. The cases with and without "normal" output node are also considered. In total there are four combinations to be tested and compared as shown in the **Table 5.1**.

Table 5.1 ANN input feature and output pattern selection

No. of ANN	Input feature	Output patterns
1	$H_2,CH_4,C_2H_2,C_2H_4,C_2H_6,CO$	Normal, Overheating, Corona, Arcing
2	$H_2,CH_4,C_2H_2,C_2H_4,C_2H_6,CO$	Overheating, Corona and Arcing
3	$H_2,CH_4,C_2H_2,C_2H_4,C_2H_6$	Normal, Overheating, Corona, Arcing
4	$H_2,CH_4,C_2H_2,C_2H_4,C_2H_6$	Overheating, Corona and Arcing

The ANN is trained using the conventional back-propagation method. Since the ANN topology has great impact on classification accuracy and convergence time during training, several network topologies are compared. Usually, one hidden-layer ANN is enough for most nonlinear mapping. However, ANNs with 1,2 and 3 hidden layers are trained and compared.

Discussed in the pervious section, as the total number of weight N increases the degrees of freedom for the classifier increase, and so does the complexity. To eliminate this variable in the evaluation process, we attempted to keep N constant for the different topologies. We first tried several networks with $N=500$, 1000, 1500 respectively and found the network with $N=1000$ had the best performance. Since it is impossible to keep the total number of weight exactly the same for different topologies. we configured the different networks with $N \cong 1100{\sim}1200$.

It is computationally prohibitive to explore all possible topologies. For each kind of network with chosen input features and output patterns, three kinds of network with one, two, and three hidden layers were generated. According to the number of training iterations and average training error, the optimal was determined. The ten-fold cross-validation of this optimal ANN was computed to verify the accuracy. Forty sample sets from different transformers were used for training and testing of the ANNs. The training and testing results of these four combinations are listed in **Table 5.2** through **Table 5.5**.

Table 5.2 ANN #1 topology selection.

ANN Topology	Training Iteration	Training Error
6-25-20-20-4	2228	0.009983
6-40-20-4	1359	0.009599
6-120-4	5000	0.015353
Ten -Fold Cross-Validation: 90%		

Table 5.3 ANN #2 topology selection.

ANN Topology	Training Iteration	Training Error
6-25-20-20-3	5000	0.011976
6-40-20-3	879	0.007286
6-120-3	5000	0.06589
Ten -Fold Cross-Validation: 86.8%		

Table 5.4 ANN #3 topology selection

ANN Topology	Training Iteration	Training Error
5-25-20-20-4	879	0.009801
5-40-20-4	663	0.009675
5-120-4	3354	0.009998
Ten -Fold Cross-Validation: 95%		

Table 5.5 ANN #4 topology selection

ANN Topology	Training Iteration	Training Error
5-25-20-20-3	2008	0.009991
5-40-20-3	1660	0.009987
5-120-3	4308	0.009988
Ten -Fold Cross-Validation: 88.7%		

The first column of the table shows the three different topologies. The 5-40-20-4 topology means a 4-layer network with 5 input nodes, 40 nodes in the first hidden layer, 20 nodes in the second hidden layer, and 4 output nodes. Other topologies are represented similarly. Comparing the accuracy, iteration number, it is shown that the two-hidden layer ANN topology is the optimal for all four cases. All ten-fold cross-validations are referred to the two-hidden layer ANNs. As can be seen, an input feature space without "CO" increases the accuracy. Keeping "normal" as an output node also increases the accuracy. The second ANN in **Table 5.4** is the optimal one (among the 12) that has 95% chance of success in classifying the three major fault types. This chosen ANN topology is shown in **Figure 5.4**.

Normal   Overheating Corona Arcing (outputs)

nodes

weighted connections

(inputs)

H_2 CH_4 C_2H_2 C_2H_4 C_2H_6
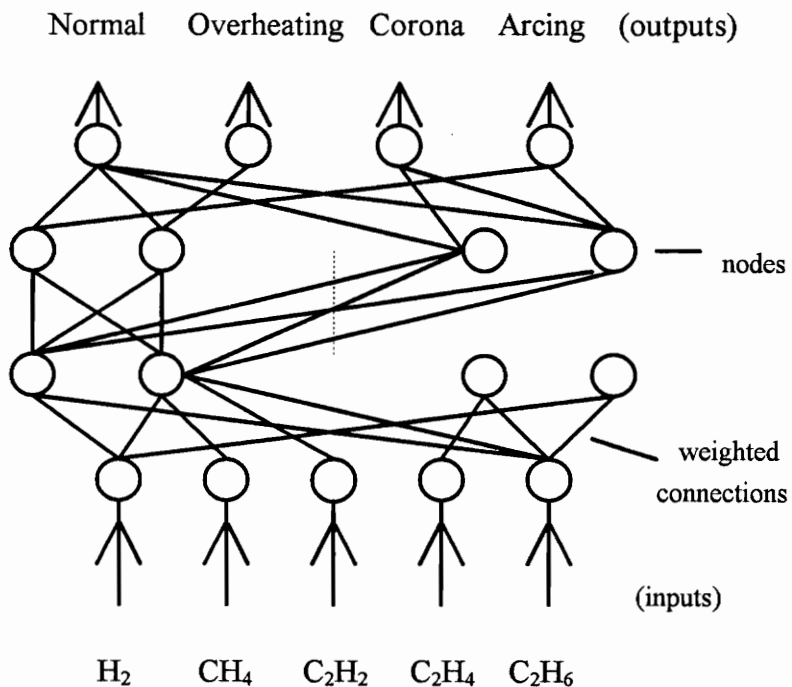
Figure 5.4 ANN for major fault type diagnosis

5.5.1 Learning Parameters

In our study, all the neural networks were trained by the delta rule with a momentum term which was introduced in the previous section. The effect of the learning parameters, learning rate (η) and momentum term (α), on the speed of training convergence was studied by experimenting with various learning rate and momentum terms.

As shown in the Equation (5-30), during the training of a back propagation network, the updating of the weight can be described by using a gradient descent rules. An obvious problem with this gradient descent algorithm is setting an appropriate learning rate. The weights are updated with an assumption that the error surface is locally linear, where "locally" is defined by the size of the leaning rate. At points of high curvature, this linearity assumption does not hold well. The learning rate is therefore usually kept low to avoid divergent behavior.

With a small learning rate, however, the neural network may converge very slowly. The concept of momentum term was introduced to compensate for this problem. Equation (5-30) is commonly modified so that a portion of the previous delta weight is fed through to the current delta weight represented in Equation (5-31). The general trends are reinforced whereas oscillatory behavior is canceled out. This allows a low learning rate η but results in a fast learning speed.

Figure 5.5 shows a representative example of convergence behavior obtained with two hidden layers and forty and twenty neurons in the two layers respectively. The

training was stopped if convergence was not reached after 5,000 iterations. each curve represents the number of iteration required to reach convergence with different values for the learning rate and momentum term. It was found that, generally, introducing a momentum term can decrease the number of iterations that are required for the neural network to reach convergence. For a small learning rate, the momentum term has a relatively large impact on the convergence speed. But as the learning rate increases, the impact decreases gradually. The final choice of the learning parameter is a relatively small learning rate (0.3) and medium momentum term (0.7).
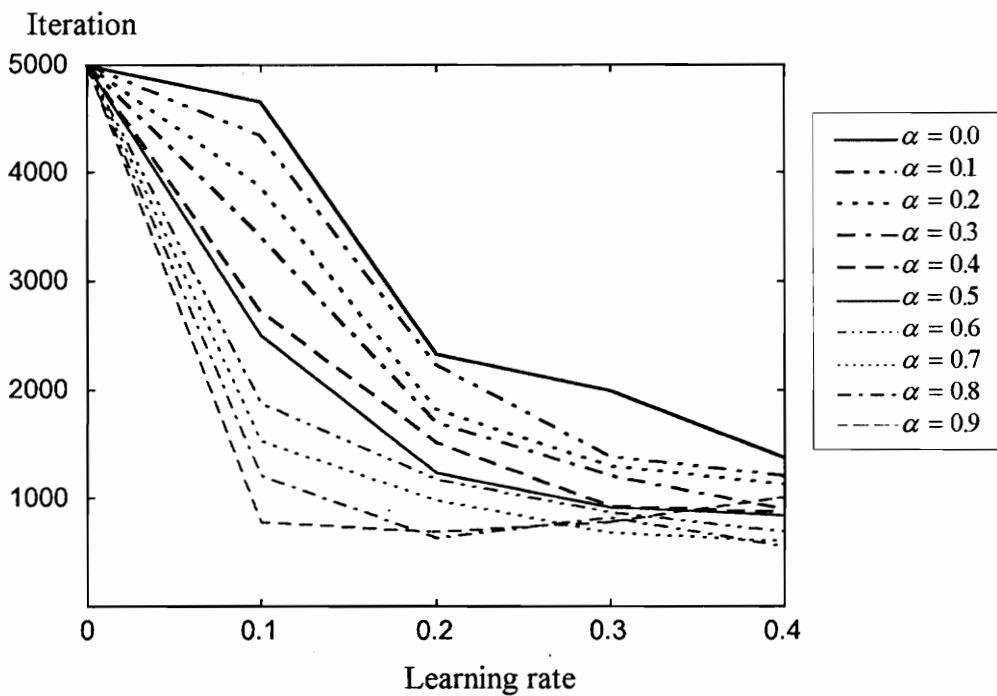
Figure 5.5 Convergence time vs. the two learning parameters

5.6 ANN for Cellulose Condition Detection

Another ANN is constructed with one output to test only the cellulose condition. Output "1" means cellulose is involved in the faults while "0" represents no cellulose involved. Although only CO and CO_2 contain the information of cellulose condition, their ratios to the hydrocarbon gases are helpful information in cellulose condition detection. Together with CO and CO_2, five hydrocarbon key gases are also used as input features. As there is only one output, training focuses on determining cellulose condition. Three topologies are tested and compared. Twenty two sample sets are used in training and testing.

According to the results in **Table 5.6**, the two hidden layer ANN is the optimal one and about 90.9% accuracy is obtained for this ANN topology as shown in **Figure 5.6**.

Table 5.6 Cellulose detection ANN topology selection.

ANN Topology	Training Iteration	Training Error
7-25-20-20-1	237	0.008876
7-40-20-1	108	0.0084589
7-120-1	211	0.009929
Ten -Fold Cross-Validation: 90.9%		

The topology of this neural network is shown in the **Figure 5.6**.

Cellulose Involved
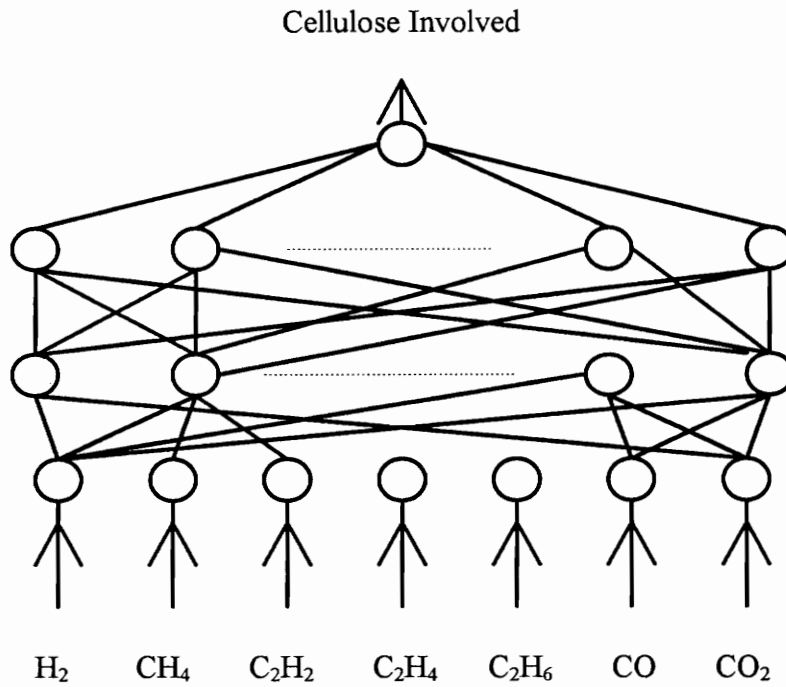
Figure 5.6 ANN for Cellulose Condition Detection

Since the CO_2 to CO ratio is an indicator of the cellulose condition, this ratio was used directly as one of the inputs in addition to the actual ppm values of CO and CO_2 in the ANN training. It is found that the rate of convergence does improve, while the overall diagnosis accuracy remains the same for the samples tested.

Compared with one step ANN approaches in [DIN95][BHA93], the training for the above two ANNs is fairly easy and the training convergence is reached without many iterations. It shows the potential of this approach as large amount of data becomes available.

The CO and CO_2 are generated from the breakdown of cellulosic materials. The amount of CO_2 to CO can be influenced to some extent by leaks during sampling and testing as CO_2 is present in detectable quantities in the atmosphere. Under some conditions carbon oxides can be generate from the oil or mineral oil containing paints. Also they are affected by the environment, temperature and etc.. Therefore, the cellulose detection is not as accurate as major fault type diagnosis. Currently, the presence of other degradation products from the cellulosic materials such as furanic compounds has been used as a confirmatory tool[GRI95]. The information about other degradation products from the cellulosic materials should be used as the input features in the ANN approach as soon as the data are available.

5.7 ANN for Trend Analysis

When a possible incipient fault condition is identified by the first time analysis, the gassing trend should be determined. Subsequent analysis provides information as to which gases are currently being generated and the rates of their generation. Even when subsequent analyses are of a routine nature, they provide the baseline from which to judge future occurrences. In the examination of trends, key gases, total combustible gas, rate of gas generation, and "fingerprints" (of normal trends) of particular types of transformers are considered.

As in the case of a first-time sample with no previous information, key gases are used to determine the type of incipient fault condition with subsequent samples, with the

difference being that only those gases actively being generated are used in the diagnosis of subsequent samples. If no significant change in gas concentration is observed for a reasonable period of time, even though values exceed norms, or ratios indicated that a problem exists, no problem may be indicated unless excessive insulation has deteriorated (which usually results in more gassing). Ratio methods may be used to aid in the analysis, but, depending upon the history of the transformer, they may not be able to distinguish between past and present conditions. For example, a transformer which has been degassed after a fault will show an altered gas pattern as the less-soluble gases are preferentially removed. Also, test results for a transformer with more than one type of fault may be difficult to interpret using ratio methods. Again, the total combustible gas content (in oil) is used to judge the relative deterioration of the insulation. It is possible that an apparent incipient fault condition will generate gases away from insulating structures and cause no harm to the transformer insulation system. However, such a condition may also mask the effects of a more serious condition developing at the same time.

The rate of gas generation helps to determine the existence of fault and also provides information as to severity of the condition and offers guidelines as to whether or not a unit should be removed from service.

Another ANN was developed to take into account the trend of gas generation. Two sets of five hydrocarbon gas ppm values and the time interval between these two sample sets are used as input features while the output patterns are still "normal," "overheating," "corona," and "arcing." The gas generation rates are indicated by these two

sets and the sampling interval. Three topologies are tested and compared. Fifty eight sample sets are selected in training and testing.

Table 5.7 ANN for major fault type detection with trend

ANN Topology	Training Iteration	Training Error
11-25-20-20-4	5000	0.256748
11-40-20-4	3953	0.046273
11-120-4	5000	0.134595
Ten -Fold Cross-Validation: 86.03%		

According to the results in **Table 5.7**, the two-hidden layer ANN is the optimal one and about 86.03% accuracy is obtained for this ANN topology as shown in the table.

From the above results, it is shown that the accuracy is about 86.03% while the accuracy is 95% as the trend is not included in the ANN.

Theoretically, the performance of the classifier should be better but it is not improved in our study. There could be some reasons causing this result. As the two set of sample data are used as input feature, the input features are doubled and thus the complexity of the system increases. The system may need more data for training and testing. However, we did not have enough data available in our study. It is suggested[GRI88] that gas generation rate 100ppm/day may indicate a fault, and it is better to use the "fingerprints" (of norm trends) which is also not available in our study.

Although the diagnosis performance is not improved by using the trend information in our study due to the lack of essential data and information, the obtained

53

performance (86.03%) is still good. We believe that the system will be improved when it is studied with enough data available.

Chapter 6 System Testing and Result Analysis

6.1 Testing Case Selection and Testing

The ANNs were constructed according to the above evaluation. The weight matrices and biases were stored as files. Testing data were inputted from a data file, then the ANNs were used to detect the faults and cellulose condition. Ten new sample sets from different transformers were tested. Nine of the cases were transformers where the results of the physical investigation were published. For case 6, the data is an average value for 200 transformers. The testing data are listed in **Table 6.1**.

Table 6.1 Testing data(ppm)

NO.	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CO	CO_2
1	280	1500	140	1200	150	94	960
2	130	98	65	56	7	110	4000
3	17000	110000	16000	89000	84000	320	430
4	300	240	140	160	14	23	160
5	48	610	ND*	10	29	1900	970
6	1565	93	0	47	34	548	8263
7	320	1370	9	1980	417	19	549
8	1400	3000	4	3500	560	1600	6800
9	1000	720	360	450	31	6	100
10	0	1	0.1	0.1	0	1	35

* ND -- Not detected

The ANNs for major fault diagnosis and cellulose degradation are implemented to diagnose these ten cases. The results are represented in **Table 6.2**.

Table 6.2 Testing Results (0 - 1 range represents the different degree and certainty of a particular diagnosis)

NO.	Inspection	Normal	Overheat	Corona	Arcing	Cellu.	ANN Diag.
1	Overheat	0.0000	0.9999	0.0000	0.0038	0.0406	Overheat
2	Arcing(p)	0.0000	0.9269	0.0003	0.9981	0.9342	A & O(p*)
3	Overheat(p)	0.0000	1.0000	0.0000	0.0000	0.9997	Overheat(p)
4	Arcing	0.0000	0.8771	0.0001	0.9998	0.2123	A & O
5	Overheat(p)	0.0000	0.9999	0.0000	0.0000	0.9998	Overheat(p)
6	Corona	0.0000	0.0419	0.9854	0.0000	0.9997	Corona(p)
7	Overheat(p)	0.0000	0.9999	0.0000	0.0000	0.5621	Overheat(p)
8	Overheat(p)	0.0000	0.9999	0.0000	0.0000	0.9768	Overheat(p)
9	Arcing	0.0000	0.0388	0.0000	0.9999	0.1807	Arcing
10	Normal	0.9865	0.1341	0.0000	0.0000	0.0650	Normal

*(p)-------- Paper or other cellulosic materials involved

6.2 Result Analysis

The results between the actual inspection of the transformer and ANN diagnosis match very well. There are two cases (2 and 4) where the ANN for major fault type diagnosis and the actual inspection disagree. These two cases where there is arcing but the ANN indicates there is arcing and overheating. The ANN can be trained with refined input data to distinguish between these two cases as the arcing degradation products are the same as the thermal degradation products but in different relative composition. There are

cases where there is overheating which causes damage to the insulation which then results in electrical discharges. It is also quite possible to have a condition which involves simultaneous overheating and electrical discharges such as from an unintentional core ground. However this does not appear to be true in these two cases.

In case 3 there is a large quantity of C_2H_2 16,000 ppm. A Quick evaluation may suggest that there is arcing in the oil as many people are sensitive to the appearance of C_2H_2. However, the relative composition of the gases indicates that the problem is thermal in nature in agreement with the ANN output and as confirmed by the internal investigation. In this case the amounts of carbon oxides is not high, however the ratio of CO_2 to CO is quite high indicating the possibility of the involvement of cellulosic materials in the arc as confirmed by inspection of the transformer.

For case 7, it is difficult to justify the ANN evaluation of the overheating of paper even though some was found on inspection of the transformer. It would appear that the ANN is picking up on the low CO_2 to CO ratio which is not always a good indicator of overheating of cellulosic materials as suggested earlier in this paper. It may be that some minimum quantity limits for these gases may be required. This can be implemented easily in the ANN computer code.

A number of the cases for which the diagnosis indicate involvement of cellulosic materials requires careful review. In case 2 there is an indication of overheating of paper both during inspection of the transformer and by ANN. There is an appreciable amount of CO_2 present. However, this is not an unusual amount and a significant portion is likely

57

from the general aging of the transformer cellulosic insulation rather than the fault. This is one of the difficulties when there is no previous data to show which gases are most actively being generated. This is similar in case 8 where it is difficult to determine without any specific previous data whether the carbon oxides generated is from the incipient-fault location or from normal aging of the cellulosic materials. In case 8 the investigation revealed loose connections resulting in hot metal and therefore it seems likely that the majority of the carbon oxides was not from the incipient-fault areas but rather from the aging of the winding insulation.

Besides the above testing examples, other 67 samples are also tested by the neural network. Among these 67 samples, 62 cases are correctly diagnosed. The accuracy of major fault type diagnosis is about 92.5% which is close to the 95% accuracy evaluated before.

In this thesis, we concentrate on the discussion of the correlation between gas in oil and the transformer incipient fault. Although, the incipient faults are the main causes of increase of gas in oil. There could be some outside causes for the concentration increase of gas in oil. The possible outside causes are listed as below:

- Welding of oil tank

- Water in the oil

- Oil filter failure

- Cooling system failure (oil pump failure)

- Transformer mount failure

- Stray magnetic flux

- Environment and temperature

- Overloading

Every time before we diagnose the transformer fault by using DGA, we should make sure that the transformer is not operating under abnormal outside conditions.

Chapter 7 Conclusions

A two-step neural-network classifier has been developed and carefully tested for transformer fault diagnosis using dissolved gas-in-oil analysis. Several feature types have been evaluated and several neural-net topologies have been considered. Patterns with different identification sensitivity are detected by different ANNs to obtain the highest accuracy for each pattern. The two-step approach makes ANN easier to train and more accurate in detecting faults.

The results of this study show that the two-ANN approach is promising for transformer fault diagnosis even with limited sample data. Theoretically, the ANN can be trained to represent any observable phenomenon if there are sufficient data available. The more complex a relationship is, the more training data are needed. The transformer fault diagnosis could be very complicated. For example, it is desirable to distinguish between faults of oil and cellulosic materials (paper), different temperatures (low, medium, and high) for overheating in oil, or low energy and high energy sustained arcing. To deal with such a complicated diagnosis problem, the available input data may not always be enough. It can be a very effective way to construct different ANNs for different pattern recognition to obtain the highest diagnosis accuracy for each pattern.

The accuracy of the diagnosis can be improved with the increase of training data. Another very important part of dissolved gas-in-oil analysis is evaluation of trends -- the increasing rates of gas generation. The trend indicates which gases are increasing and at

what rate of generation. For example, a transformer with a moderate amount of gases can be considered perfectly normal until 7 ppm of acetylene suddenly is detected. The 7 ppm of acetylene may represent a very small fraction of the total combustible gas, but it could be the most important information contained on the data. An ANN taking into account the gas generation trend is also developed and works well in this study. It shows that once enough information is available, the diagnosis can be made more reliable. Due to the limited trend data available, this work will be done at a later stage.

Bibliography

[AND76] R. Andersson, U. Roderick, V. Jaakkola, N. Ostmann, "The transfer of fault gases on transformers and its effect upon the interpretation of gas analysis data." CIGRE 1976 Report 12-02.

[BHA93] S. K. Bhattacharyya, R. E. Smith, and T. A. Haskew, "A neural network approach to transformer fault diagnosis using dissolved gas analysis data", NAPS, 1993.

[CIG75] CIGRE 15-10. "Detection of and Research for the Characteristics of an Incipient Fault from Analysis of Dissolved gases in the Oil of an Insulation." Electra 42 (October 1975) pp31-52.

[DIN93] X. Ding, Y. Liu, " Predicting Power Transformer Resonant Frequencies Using ANN", NAPS, 1993.

[DIN95] X. Ding, E. Yao, Y. Liu, P. J. Griffin, "ANN based transformer fault diagnosis using gas-in-oil analysis", 57th American Power Conference, April 1995, Chicago IL.

[DUV74] M. Duval. "Fault gases in oil-filled breathing EHV power transformers. The interpretation of gas analysis data." IEEE PES Conference paper No. C74 476-8, 1974.

[ERD88] Criteria for the Interpretation of Data for Dissolved Gases in Oil Transformers (A review), Electrical Insulating Oils, STP 998, H. G. Erdman, Ed., American Society for testing and Materials, Philadelphia, 1988, pp89-106.

[FAL70] B. Fallou, F. Viale, I. Davies, R. R. Rogers, E. Dornenburg. "Application of Physical-Chemical Methods of Analysis to the study of deterioration in the insulation of electrical apparatus." CIGRE 1970 Report 15-07.

[GRI88] P. J. Griffin, "Criteria for interpretation of data for dissolved gases in oil transformer (A Review)", Electrical Insulating oils, STP998, American Society for Testing and Materials, Philadelphia, 1988, pp89-106.

[GRI95] P. J. Griffin, and L. R. Lewand, "A Practical Guide for Evaluating the Condition of Cellulosic Insulation in Transformers," Doble Client Conference Proceedings, 1995, Sec 5 - 6.1.

[HAL73] W. D. Halstead, "A thermodynamic Assessment of the Formation of Gaseous Hydrocarbons in Faulty Transformers". J. Inst. Petroleum 59 (Sept. 1973) 569 pp239-241.

[HOW92] Howard Demuth and Mark Beale, "Neural Networks Toolbox," The Math Works, Inc., 1992.

[INO90] Y. Inoue, K. Suganuma, M. Kamba, and M. Kikkawa, "Development of oil-dissolved hydrogen gas detector for diagnosis of transformers," IEEE Trans. on Power Delivery, Vol.5, No.1, January 1990.

[LIN93] C. E. Lin, J. M. Ling, and C. L. Huang, "An Expert System for Transformer Fault Diagnosis Using, Dissolved Gas Analysis Analysis," IEEE Trans. on Power Delivery, Vol.8, NO.1, Jan. 1993, pp231-238.

[NEK95] R. Nekovei and Y. Sun, "Back-Propagation Network and its Configuration for Blood Vessel Detection in Angiograms," IEEE Trans. on Neural Networks, Vol.6, No. 1, January 1995, pp.64-72.

[PAO89] Y. H. Pao, "Adaptive Pattern Recognition and Neural Network", Addison-Wesley publish company, Inc., New York, NY, 1989.

[ROG78] R. R. Rogers, "IEEE and IEC Codes to Interpret Incipient-Faults in Transformers Using Gas in Oil Analysis", IEEE Trans. Electr. Insul, Vol. EI-13 No.5. October, 1978, pp349-353.

[ROG78] R. R. Rogers, "IEEE And IEC Codes To Interpret Incipient Faults In Transformers, using Gas In Oil Analysis," IEE Trans. Electr. Insul, Vol EI-13 No 5, October 1978.

[TOM93] K. Tomsovic, M. Tapper, T. Ingvarsson, "A Fuzzy Information Approach to Integrating Different Transformer Diagnostic Methods," IEEE Transaction on Power Delivery, Vol.8, No.5 July,1993, pp1638-1646.

[WEI91] S. M. Weiss, and C. A. Kulikowski, "Computer Systems, That Learn," San Mateo: Morgan Kaufmann Publishers, Inc. 1991.

[ZHA96] Y. Zhang, X. Ding, Y. Liu, P. J. Griffin, "An Artificial Neural Network Approach to Transformer Fault Diagnosis," IEEE Paper 96 WM225-3 PWRD presented at the IEEE/PES 1996 Winter Meeting, Baltimore, to appear in IEEE Trans. Power Delivery, 1996

Appendix

A.1 Training Data

The following three tables list the data used to train the ANNs. These three ANNs are ANN for major fault diagnosis, ANN for major fault diagnosis with trend and ANN for cellulose condition detection .

The meanings of notations used in the tables are as below:

      N --- Normal

      O --- Overheating

      C --- Corona

      A --- Arcing

Table A.1 Training data for major fault type diagnosis

NO.	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	Fault
1	0	1.0	0.1	0.1	0	N
2	35.0	25.0	22.0	23.0	0	A
3	10.0	4.0	6.0	33.0	3.0	N
4	370	28	0	49	19	C
5	168.0	1353	63	3281	581	O
6	770	1420.0	3.0	1452	401.0	O
7	57.0	13.0	12.0	11.0	0.1	A
8	23.0	506.0	213.0	7836.0	2229.0	O
9	98.0	252.0	22.0	646.0	95.0	O
10	42.0	62.0	73.0	63.0	5.0	A
11	557.0	3541.0	7.0	2928.0	521.0	O
12	851.0	2802.0	10.0	3521.0	772.0	O
13	100.0	170.0	0	34.0	595.0	O
14	97.0	405.0	34.0	503.0	190.0	O
15	23.0	12.0	61.0	10.0	12.0	A

16	144.0	754.0	1.0	562.0	250.0	O
17	240.0	676.0	4.0	818.0	200.0	O
18	0	434.0	0	387.0	226.0	O
19	2844.0	8518.0	39.0	10196.0	4422.0	O
20	117.0	357.0	4.0	468.0	92.0	O
21	335.0	95.0	203.0	98.0	11.0	A&O
22	38.0	43.0	8.0	67.0	11.0	A
23	93.0	20.0	15.0	9.0	2.0	A&O
24	1410.0	4450.0	1480.0	8490.0	1440.0	A&O
25	612.0	810.0	520.0	325.0	170.0	A
26	940.0	264.0	272.0	199.0	16.0	A
27	8.0	631.0	39.0	2020.0	254.0	O
28	1770.0	3630.0	78.0	8480.0	1070.0	O
29	218000.0	1910.0	5.0	178.0	262.0	C
30	5.0	217.0	6.0	523.0	69.0	O
31	363.0	198.0	444.0	269.0	41.0	A
32	62.0	613.0	3.0	1180.0	202.0	O
33	7220.0	7200.0	1850.0	7180.0	643.0	A&O
34	509.0	3000.0	215.0	5380.0	1310.0	O
35	9817.0	36962.0	2624.0	62815.0	11608.0	O
36	03.0	1768.0	2.0	1302.0	200.0	O
37	1000.0	4300.0	24.0	5400.0	1100.0	O
38	240.0	28.0	85.0	26.0	6.0	A
39	0	2.0	0	8.0	2.0	O
40	3510.0	45.0	0	92.0	37.0	C

Table A.2 Training data for major fault diagnosis with trend

Sample No.	Late sample set					Early sample set					Interval (month)	Fault type
	H$_2$	CH$_4$	C$_2$H$_2$	C$_2$H$_4$	C$_2$H$_6$	H$_2$	CH$_4$	C$_2$H$_2$	C$_2$H$_4$	C$_2$H$_6$		
1	0	1.0	0.1	0.1	0	0	1.0	0.1	0.1	0	1	N
2	35.0	25.0	22.0	23.0	0	35.0	25.0	22.0	23.0	0	1	N
3	370	28.0	0	49.0	19.0	370.0	28.0	0	49.0	19.0	1	C
4	168	1353	63.0	3281	581.0	168.0	1353	63.0	3281	581.0	1	O
5	770	1420	3.0	1452	401.0	770.0	1420	3.0	1452	401.0	1	O
6	57.0	13.0	12.0	11.0	0.1	57.0	13.0	12.0	11.0	0.1	1	A
7	23.0	506	213	7836	2229	23.0	506.0	213.0	7836	2229	1	O
8	98.0	252	22.0	646.0	95.0	98.0	252.0	22.0	646.0	95.0	1	O
9	42.0	62.0	73.0	63.0	5.0	42.0	62.0	73.0	63.0	5.0	1	A
10	557	3541	7.0	2928	521.0	557.0	3541	7.0	2928	521.0	1	O
11	851	2802	10.0	3521	772.0	851.0	2802	10.0	3521	772.0	1	O
12	100	170	0	34.0	595.0	100.0	170.0	0	34.0	595.0	1	O
13	97.0	405	34.0	503.0	190.0	97.0	405.0	34.0	503.0	190.0	1	O
14	23.0	12.0	61.0	10.0	12.0	23.0	12.0	61.0	10.0	12.0	1	A
15	144	754	1.0	562.0	250.0	144.0	754.0	1.0	562.0	250.0	1	O
16	240	676	4.0	818.0	200.0	240.0	676.0	4.0	818.0	200.0	1	O
17	0	434	0	387.0	226.0	0	434.0	0	387.0	226.0	1	O
18	2844	8518	39.0	10196	4422	2844	8518	39.0	10196	4422	1	O
19	1770	3630	78	8480	1070	3450	3240	62	7730	992	1.3	O
20	335	95	203	98	11	70	19	0	40	6	12	O&A
21	38	43	8	67	11	44	49	10	78	13	3	A
22	93	20	15	9	2	19	12	0	5	1	14	O&A
23	1410	4450	1480	8490	1440	0	2	0	7	4	32	O&A
24	612	810	520	325.0	170.0	612.0	810.0	520.0	325.0	170.0	1	A

#	940	264	272	199.0	16.0	940.0	264.0	272.0	199.0	16.0		
25	940	264	272	199.0	16.0	940.0	264.0	272.0	199.0	16.0	1	A
26	8	631	39	2020	254	8	715	50	2450	315	2	O
27	1770	3630	78.0	8480	1070	3450	3240	62.0	7730	992.0	1.3	O
28	21800	1910	5.0	178.0	262.0	21800	1910	5.0	178.0	262.0	1	C
29	5.0	217.0	6.0	523.0	69.0	5.0	217.0	6.0	523.0	69.0	1	O
30	363.0	198.0	444	269.0	41.0	363.0	198.0	444.0	269.0	41.0	1	A
31	62	613	3	1180	202	75	270	3	510	64	44	O
32	7220	7200	1850	7180	643	0	0	0	0	0	2	O&A
33	509.0	3000	215	5380	1310	509.0	3000	215.0	5380	1310	1	O
34	9817	36963	2624	62815	11608	9817	36963	2624	62815	11608	1	O
35	403.0	1768	2.0	1302	200.0	403.0	1768	2.0	1302	200.0	1	O
36	1000	4300	24.0	5400	1100	1000	4300	24.0	5400	1100	1	O
37	240.0	28.0	85.0	26.0	6.0	240.0	28.0	85.0	26.0	6.0	1	A
38	0	2.0	0	8.0	2.0	0	2.0	0	8.0	2.0	1	N
39	3510	45	0	92	37	68	11	0	7	7	42	C
40	86.3	188.1	2.6	397.6	71.4	1.0	11.2	0	56	4.9	15	O
41	356	1206.4	1.1	1659	354.4	4	6.3	0.8	18.8	3.9	7	O
42	81	7.3	111	6.3	26.4	31	6.6	67	4.7	19	1	O&A
43	44	12.2	17.4	3.2	3.4	16.7	2	1.8	7.5	1.4	0.5	O
44	81.2	135.4	1.6	284	29.2	0	3.2	0.5	29.7	7.5	2.5	O
45	0	1437	2904	2378	263	0	2.13	0	2.65	3.29	0.7	A
46	54.2	211.1	2.43	724.8	140.4	0	17.8	0	18.2	8.26	1	O
47	144.7	202	10.2	486.7	77.2	0	7.2	2.58	4.06	7.32	3	A
48	442.9	85.21	174	103	9.5	20.83	15.77	0.16	18.71	3.57	1	O
49	68	100	0	177	70	41	58	0	124	50	0.1	A&O
50	77.9	153.2	3.75	337	96.4	14	30.2	3.14	158.4	29	7	O
51	21.75	55.13	15.51	15.58	12.62	0	38.29	0	5.28	9.44	0.3	A
52	20.99	78.95	0.37	58.24	20.21	10.02	91.34	0	44.4	14.77	0.3	A

53	353.5	13.53	187.3	142.6	10.23	6.05	9.42	0.74	5.27	2.21	0.3	A
54	336	419	21	1074	105	0	1	0	12	1	6	A&O
55	61.6	1032	0	620.0	144.0	49.3	319.4	0	449.5	169.4	5	O
56	90	124	5.9	259	26	79	27	2.3	45	6.8	2.5	O&A
57	0	3.4	3.78	1.6	0	23.53	2.47	4.36	0	0	1	A
58	11	19	3	103	3	160	223	11.0	495	45	2	O&A

Table A.3 Training data for cellulose condition detection

NO.	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CO	CO_2	Cellu.
1	10.0	4.0	6.0	33.0	3.0	39.0	2907.0	No
2	335.0	95.0	203.0	98.0	11.0	440.0	2170.0	Yes
3	38.0	43.0	8.0	67.0	11.0	37.0	519.0	No
4	93.0	20.0	15.0	9.0	2.0	276.0	5540.0	Yes
5	1410.0	4450.0	1480.0	8490.0	1440.0	341.0	376.0	Yes
6	612.0	810.0	520.0	325.0	170.0	713.0	2760.0	Yes
7	940.0	264.0	272.0	199.0	16.0	45.0	1290.0	No
8	8.0	631.0	39.0	2020.0	254.0	23.0	1390.0	No
9	1770.0	3630.0	78.0	8480.0	1070.0	832.0	7940.0	Yes
10	21800	1910.0	5.0	178.0	262.0	289.0	787.0	Yes
11	5.0	217.0	6.0	523.0	69.0	151.0	2610.0	No
12	363.0	198.0	444.0	269.0	41.0	208.0	657.0	Yes
13	62.0	613.0	3.0	1180.0	202.0	638.0	8910.0	Yes
14	7220.0	7200.0	1850.0	7180.0	643.0	4670.0	2290.0	Yes
15	509.0	3000.0	215.0	53800	1310.0	292.0	4720.0	Yes
16	0	2.0	0	8.0	2.0	4.0	186.0	No
17	3510.0	45.0	0	92.0	37.0	344.0	8170.0	Yes
18	1873.0	934.0	1547.0	800.0	78.0	576.0	2440.0	No
19	158.0	77.0	0	62.0	17.0	1050.0	10894.0	Yes
20	420.0	4400.0	5.0	3900.0	950.0	42.0	1300.0	No
21	42.0	132.0	0	36.0	54.0	219.0	2641.0	Yes
22	32.0	27.0	0	15.0	8.0	343.0	2791.0	Yes

A.2 Testing Data

The following 67 testing data are obtained from various references. The last column list the real fault types. ANN diagnosis results are given in the parenthesis( ), if they do not match the real results.
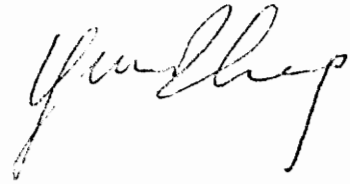
Table A.4 Testing data

NO.	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CO	CO_2	Fault
1	685	102	367	170	9	211	1736	A
2	64	250	23	846	145	265	3887	O
3	128	419	0.35	614.1	269.4	833	1423	O
4	236	410.2	3.5	817.3	159	200	893	O
5	73	520	6	1200	140	410	550	O
6	4	29	0.3	64	8	ND	ND	O
7	42	97	0	600	157	213	2120	O
8	5	11	0	70	18	ND	ND	O
9	160	223	11	495	142	233	3590	O
10	17	29	1.4	64	6	ND	ND	O
11	385	381	1	596	142	233	3590	O(A)
12	26	34	0.1	53	13	ND	ND	O
13	766	993	4	665	116	29	230	O
14	30	56	0.2	37	7	ND	ND	O
15	246	1060	24	1280	216	308	1860	A
16	9	41	1	50	8	ND	ND	O
17	39	1660	0	2800	890	69	2540	O
18	0.6	31	0	52	17	ND	ND	O(N)
19	16	237	0	470	92	157	1620	O
20	2	30	0	59	12	ND	ND	O
21	56	286	7	928	96	60	2280	O
22	4	22	0.5	70	7	ND	ND	O(A)
23	15	125	7	574	29	141	3140	O
24	2	17	0.1	78	4	ND	ND	O
25	109.59	133.96	23.43	361.00	39.30	760.7	7870.6	A&O
26	36.2	12.6	0	15.5	4.4	1981	14746	A&O
27	32.7	14.6	0	15.8	2.8	1082	11667	A&O
28	41	10	1	13	7	481	5061	A
29	692	605	2833	1437	78	1245	5918	A
30	336	419	21	1074	105	99	1091	O

31	73	110	0	190	41	290	1200	A&O
32	77	120	0.8	210	44	310	1200	A&O
33	95	130	1.2	220	55	360	1200	A&O
34	76	140	1.1	230	64	340	1300	A&O
35	1.0	11.2	0	56	4.9	110	2334	N
36	86.3	188.1	2.6	397.6	71.4	557	3821	O
37	365	1206.4	1.1	1658.6	354.4	228	1670	O
38	23.53	2.47	4.36	0	0	46.1	654.8	A
39	0	3.4	3.78	1.6	0	96.4	882.1	N
40	81	7.3	111	6.3	26.4	290	1785	O
41	59	7.1	71	4.5	19	243	972	A
42	31	6.6	67	4.7	19	329	1898	A
43	16.7	2	1.8	7.5	1.4	126.8	3436	N
44	44	12.2	17.4	3.2	3.4	206	4488	A
45	81.2	135.4	1.6	284	29.2	363.5	5928	A
46	466.5	148.8	511.5	266.4	13.1	117.7	309	A
47	41.9	30.1	177.6	139.6	9.8	23.3	746.5	A&O
48	54.2	211.1	2.43	724.8	140.4	1380.6	8709	A
49	0	7.2	2.58	4.06	7.32	442	1515.5	A
50	144.7	202	10.2	486.7	77.2	605	2312	A
51	20.83	15.77	0.16	18.71	3.57	368.14	3817.7	A
52	442.94	85.21	173.92	102.93	9.5	1002.4	6512.3	O
53	68	100	0	177	70	ND	ND	A&O
54	14	30.2	3.14	158.4	29	274	2520	A&O
55	77.9	153.2	3.75	337	96.4	1048	7626	A&O
56	64.5	12.7	17.64	9.1	1.44	ND	ND	C(A)
57	135	17.9	96.6	17.8	1.6	ND	ND	C
58	1267	1473	2904	2378	263	1692	3702	O
59	11949	7683	16108	9475	576	19155	2319	A&O
60	44.35	17.36	10.317	23.33	3.67	488.58	4965.8	O
61	442.94	85.21	173.92	102.93	9.50	1002.4	6512.3	A
62	6.05	9.42	0.74	5.27	2.21	259.7	1548.2	N
63	21.75	55.13	15.51	15.87	12.62	365	3675	A&O
64	20.99	78.95	0.37	58.24	20.21	50.25	2367	A
65	336	419	21	1074	105	99	1091	A
66	0	1.7	0.51	5.59	1.4	3	480	A&O
67	0	1.2	0.20	2.34	0.69	22	500	O(A)

*ND --- Not detected

Vita

Yuwen Zhang was born in Shanghai, the People's Republic of China on January 28, 1964. He attended Shanghai Jiao Tong University, P. R. China, in the Fall of 1981, and obtained both his B.S. and M.S. degrees in Electrical Engineering in July 1985 and March 1988 respectively. After having taught in Shanghai Jiao Tong University for six years, he attended Virginia Polytechnic Institute and State University(Virginia Tech) in the Bradley Department of Electrical Engineering in the Spring of 1994. His graduate study at Virginia Tech has included electrical device fault detection and diagnosis, logic/circuit design, and communications.