Classification of Faults and Switching Events by Inductive Reasoning and Expert System Methodology

C. J. Kim
Student Member

B. Don Russell
Senior Member

Texas A&M University
College Station, Texas 77843

ABSTRACTS

Under certain conditions, one electrical parameter (independent variable) is not enough to detect high impedance faults on certain surface conditions. These faults do not draw sufficient current for detection and may draw less current than similar faults on other soil surfaces. Moreover, because every electrical detection parameter displays characteristics of randomness, it is difficult to assign a probability that a given event is a high impedance fault versus a switching event. It has been shown that detection by induction laws can improve the classification of faults and switching events. The second and third laws of induction are utilized with a minimum entropy method. Setting detection threshold values using induction methods is also proposed. The methods presented in this paper are taken from ongoing research in high impedance fault detection. While the techniques have not been reduced to practice or field tested, they hold promise for future improvements in the relaying of high impedance faults.

Keywords: high impedance fault, expert system, induction theory, minimum entropy, pattern classification, learning.

INTRODUCTION

Conventional expert systems approach final decisions by assigning a basic probability to each event and then calculating belief functions based on this probability assignment[1]. By the combination rule, it is possible to achieve a combinational belief function. This concept holds some promise for detecting high impedance faults and distinguishing these faults from normal system activity and switching events on a distribution feeder.

A basic problem with this method consists of the initial assignment of probabilities for each event measured on the distribution feeder. As has been shown in previous work[2], there are numerous electrical parameters which give an indication of high impedance faults, but which also may be active for switching and other normal events. The amount of data available is often not sufficient to draw a deterministic conclusion. A trained observer can still decide whether a waveform represents a fault or not, but extracting sufficient indicators from this expertise is quite difficult. An approach which has been used recently is to set threshold values for fault detection in an arbitrary, but expert manner. For example, earlier algorithms have used the "noise" from arcing faults as an indicator of the presence of a high impedance fault versus a switching event on the feeder. Using several electrical parameters, we can monitor the relative increase in amplitude over a particular window of investigation (e.g. 30 cycles). Based on the level of the amplitude over this window, the status of the system, either fault or not, is decided.

While this system works under certain select conditions, high impedance faults on unusual surfaces conditions such as concrete or asphalt are quite difficult to separate from switching events based on the observed behavior of the waveform. It has been shown that using one variable such as amplitude increase or randomness is insufficient for full classification. The use of several detection variables is therefore deemed necessary for secure separation of faults from normal activity.

Even though we may use several variables, it is quite difficult to assign probability functions to each variable as described above. Our objective is to utilize all the information available to us on each variable while remaining within the practical limits of software implementation. After considerable investigation, the concept of using induction with minimum entropy to classify and recognize fault patterns was adopted [3,4].

METHODOLOGY

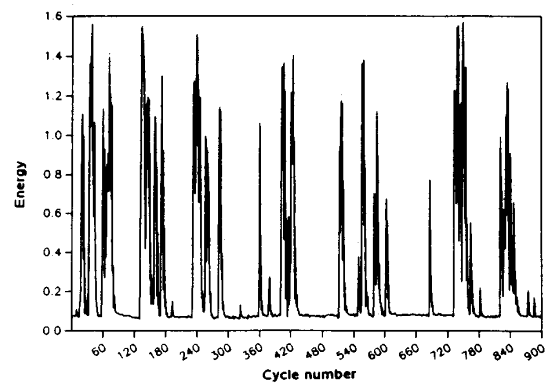
High impedance faults are characterized by a high degree of randomness in the magnitude changes and burst durations of the harmonic currents[2]. Previous research at Texas A&M University has developed several algorithms to take advantage of the increase in "energy" level and the degree of "randomness" associated with arcing, high impedance faults [5,6]. In these algorithms, one compares electrical parameters measured during staged faults to normal states to develop expertise as to the different behavior of these parameters. Threshold values for detection are based on somewhat arbitrary means. Only after developing significant insight and expertise from reviewing numerous fault scenarios can these thresholds be set by intuition. The various parameters which have been used for detection include various harmonics(even, odd, or "in-between") and high frequency components(2 KHz and above). Under certain conditions, one parameter may be sufficient to classify the fault, but in other cases, several parameters may prove insufficient. These differences are exemplified by the following facts. Figure 1(a) shows the cycle "energy" level of arcing fault on wet soil. Figure 1(b) shows the cycle "energy" level of arcing fault on grass. Figure 1(c) shows the cycle "energy" level of an air switch switching on and off. The parameter used is a composite of all the odd harmonics up to 2KHz which are filtered from the fault waveform through the 60Hz notch filter, then the comb type filter, and sampled with a rate of 7200Hz. Here the "energy" means the summation of the magnitude of squared sample values over one 60 Hz cycle.

When we observe high impedance fault waveforms for faults on asphalt and grass, the current amplitude changes and ran-

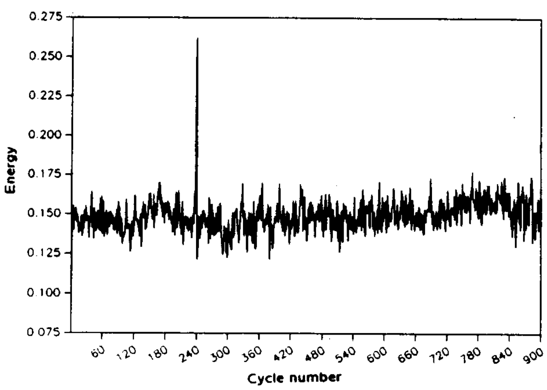
89 WM 058-9 PWRD A paper recommended and approved by the IEEE Power System Relaying Committee of the IEEE Power Engineering Society for presentation at the IEEE/PES 1989 Winter Meeting, New York, New York, January 29 - February - 3, 1989. Manuscript submitted September 1, 1988; made available for printing December 14, 1988.

domness are at a reduced level compared to certain faults on bare soil. Separation of these faults from switching events can be difficult using only one parameter.

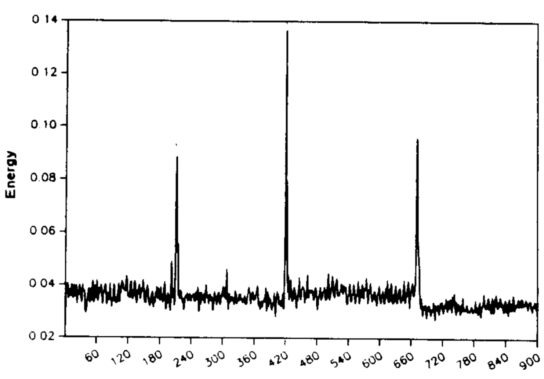
To resolve this problem, we turn to the concept of entropy. Entropy is a measure of the disorder in classifying data. Entropy minimization is an ordering principle by which we determine which past events are more like our future events in ways that are sufficient for predicting the outcome. If one wishes to classify a set of samples into a more ordered state, he attempts to lower the entropy. Therefore, by minimizing entropy, we can induce rules for classification. The following is an introduction into the use of induction laws as they apply to pattern recognition for the purposes of fault classification.



(a) Fault on Wet Soil



(b) Fault on Wet Grass



(c) Air Switch On and Off

Figure 1. Waveforms of Three different Sample events

INDUCTION THEORY

The essential principles of induction have been known for centuries. The first is Bernoulli's Principle of Insufficient Reason: If there is insufficient reason to suppose that two probabilities are not equal, then they are to be treated as equal. The second is Poisson's Frequency Limit Hypothesis: As the number of observations of an event increases without bound, the frequency of an outcome approaches the probability of the outcome. The third is Ockhams's Razor: The induced rule is the simplest rule consistent with all available information.

Three laws of induction are summarized as below[3]:

1) Given a set of irreducible outcomes of trial, the induced probabilities are those probabilities consistent with all available information which minimize the entropy of the set.
2) The induced probability of a set of independent observation is proportional to the probability density of the induced probability of a single observation.
3) The induced rule is that rule consistent with all available information for which the entropy is minimum

We do not need to assign probability factors if we can separate faults from switching events easily and formally. The third law is appropriate for detection and we might use the second law to calculate the reliability of the rule and mean probability of each step of separation. From the second law, the following is derived[3].

$$<p>=\frac{x+1}{n+2}$$

where,

$<p>$ is a mean probability(when there are only two classes).

n is the number of observations.

x is the number of observations classified as True.

Here, either True or False can be assigned to indicate Fault. If one is chosen, the other indicates switching events.

More general mean probability is is given by

$$<p>=\frac{x+t}{n+t+f}$$

where:

t is the number of distinguishable True states.

f is the number of distinguishable False states.

n is the number of observations.

x is the number of observations classified as True.

The classical problem of the third law of induction is the problem of pattern recognition, that is, classification. In pattern recognition we completely disregard the probability aspects of the problem and simply ask whether it is True or False.

For classification we need as many independent variables as possible. Initially, we find values of each variable at each sample and then make 1/0 tables with the data. For example, for the total number of burst cycles in a given window length, there are many cases in which the value varies from 0 to 30. So, one usually chooses the median as a border line of 1/0 logic. The reason for using the median is to set the border line at the medium value of all values. Median is not always the best for classification, so another method, threshold value, will be treated later.

So if the median is 5 for example, and if a sample shows 6 for this independent variable, the variable is assigned a "1". A value of 4 would correspond to "0". With this 1/0 table, the next step taken is to find rules to recognize the True and False pattern.

For seven variables, there are seven digit numbers and it is known which numbers are in the class of True and False. The procedure is to select one from the possible rules. This rule should have minimum entropy. There are many possible rules and it takes much time to derive all the possible rules and corresponding entropies and determine which one has minimum entropy. A simplified approach is now introduced. At each step, find one variable which has maximum weight or importance and then separate samples into two classes by this variable. All the values are converted to 1 or 0 logic; therefore, one variable is just a digit in a multi-digit number. Finally, calculating the reliability of the induced rules tells how reliable the induced rule is.

GENERAL APPROACH TO PATTERN CLASSIFICATION

The entropy on a set of possible outcomes of a trial where one and only one outcome is True is defined as:

$$S = -k \sum_{i=1}^{N} p_i \ln p_i$$

In other words, the entropy is the expected value of the information. Mathematically, the information contained in an event that has a probability p of occurring is given by the formula $I = -k \ln p$, where k is any positive constant, and I is the information contained. What this means is that the more unlikely the occurrence of the event, the greater the amount of information gained by observing the event. In order to understand the relationship between entropy and information in information theory, it is necessary to first recall what is meant by expected value. Suppose that, on a given trial there are several possible outcome events , one and only one of which will occur. Suppose that a value is assigned to each of the possible outcome events (e.g., the amount of money one would win should that particular outcome occur). The expected value of the trial is then simply the sum of the weighted values of the various possible outcomes, each outcome being weighted according to its probability of occurrence. Suppose now that the value assigned to each of the possible outcomes of events is its information value, which is of course uniquely determined by its probability according to the above formula. Then, if the information value of each of the possible outcome events is weighted by its probability of occurrence and these weighted information values are summed, the expected information value of the trial is obtained. This is what is called the entropy of the set of the possible outcomes of the trial.

On the other hand, the entropy of a rule is defined as below:

$$S = -k \sum_{i=1}^{m} x_i p_i \ln p_i$$

where,
m is the number of total steps
i is the step
x_i is the number of samples of either True or False at step i
p_i is a mean probability of the i^{th} step for either True or False
and k is a constant.

The third law of induction which is typical in pattern recognition says that the entropy of a rule should be minimized to have a simple and reliable rule. To find the minimum entropy and its corresponding rule, all the possible combinations of steps and of rules are to be investigated. The number of possible n digit numbers is $N = 2^n$. A rule which separates these N numbers into two classes is desired. There are N^2 ways of separating N numbers into two classes. There are only m samples. Then the number of available patterns are 2^m. This means, if there are 7-digit numbers and only 31 sample patterns, then there are 2^{31} ways of separation.

For simplifying the steps in rule induction, some easily derivable relations and tips for entropy minimization are

1) the higher p_i is, the smaller S is.
2) the bigger x_i is, the higher p_i is.
3) the bigger the index of digit n is, the bigger x_n is.

Also, the digit index is defined as the rating of separation into two classes in each digit for both True and False classes.

For example, suppose there are two classes of True and False and each class has four, 3-digit numbers as below:

TRUE	FALSE
101	001
010	110
011	111
000	100

If we consider the first digit index, and choose 0xx as True(x means don't care), then we have one wrong separation in True and another one in False. For the second digit, if we choose either x0x as True or x1x as True, we have two wrong separations at both classes. So, apparently using the first digit is better than using the second digit.

Next it is necessary to find an index of a digit to find out which digit is most important to separate numbers into two classes.

First, count the number of 1's in True and count the number of 0's in False, and divide each number by the number of samples in each class. Then the digit counts, d_n's are as below:

TRUE			FALSE		
d_1	d_2	d_3	d_1	d_2	d_3
0.25	0.5	0.5	0.25	0.5	0.5

Then, adding together digit by digit, the result is:

d_1	d_2	d_3
0.5	1.0	1.0

If the value of d_n is closer to 1, that digit is not important to separate. That is, 1's and 0's have the same weight(number, frequency, or importance) on both sides. If the value of d_n is away from 1, there are less 1's or 0's in one class, so the index of the digit n, I_n, is defined here using d_n. The maximum value of I_n is 1.

$$I_n = |d_n - 1|$$

Therefore, from the above example,

$I_1 = 0.5$, I_2=0.0, and I_3 =0.0, so the first digit has maximum index of digit. This digit can be used to separate two classes if only one digit is used to classify.

The following is the simplified version of entropy minimization:

1) Find maximum index of digit n. Then separate two classes by either 1 for True or 0 for False.
2) Eliminate those samples which are subset of the above step.
3) Find the maximum index of the digit from remainders.
4) Do steps 1) - 3) until two classes are empty.

THRESHOLD VALUE

In a discrete system, it is easy to assign 1's and 0's, but in continuous system, certain values must be set to divide the sample events into 1's and 0's. This is called a threshold value. If the threshold value is changed, the 1/0 table is changed and so are the steps, rule and entropy value. How to set a threshold value to separate two classes efficiently and simply is the point of the next discussion.

The idea is simple: find a threshold value of each independent variable which makes a minimum entropy for that independent variable. In this case, the problem is still a classification system. The only difference is, this case is inside of an independent variable, but the previous one(rule induction) is the classification with all the independent variables. So, the idea of minimum entropy of the induced rule is still used here in threshold value calculation. Thus, finding a threshold value is a kind of finding rule for an independent variable. Assume that a threshold value for an independent variable x in the range of $X1$ to $X2$ is being sought. Then for this independent variable x only, the entropy equations are written as below[2]:

$$S(x) = p(x)S_p(x) + q(x)S_q(x)$$
$$S_p(x) = -\sum_{k=1}^{m} p_k(x)\ln p_k(x)$$
$$S_q(x) = -\sum_{k=1}^{m} q_k(x)\ln q_k(x)$$

where,

$S(x)$: entropy of an independent variable x

$S_p(x)$: entropy of an independent variable x in the region $X1$ to $X1+x$

$S_q(x)$: entropy of an independent variable x in the region $X1+x$ to $X2$

$p_k(x)$: probability of the k^{th} outcome class given that the independent variable value is in the region $X1$ to $X1+x$.

m: number of outcome classes. It is 2 for True and False

$q_k(x)$: probability of the k^{th} outcome class given that the independent variable value is in the region $X1+x$ to $X2$.

Relatively unbiased estimates for $p_k(x)$ and $p(x)$ are:

$p_k(x) = n_k(x)/n(x)$, $\quad p(x) = n(x)/n$

where,

$n_k(x)$: the number of samples located between $X1$ and $X1+x$ in the k^{th} outcome class.

$n(x)$: total number in this region in all the classes.

n: total number of samples.

Now x can be chosen by various manipulation. This x can be any value between $X1$ and $X2$. Here we make steps in the value of x: i.e., x_i's. If $S(x_i)$ is minimum, then $X1 + x_i$ is the threshold value. So if each digit has its minimum entropy at the threshold value, then the total entropy for rule is also minimized.

The following is a simplified and coarse example for finding a threshold value. Here we do not use even steps but randomly choose x_i, for example purposes, for the digit n(even steps are shown in the Application Example section). Here we have only two classes, so k in the above equations has values 1 and 2 or T(for True) and F(for False). Therefore m equals 2.

TRUE : 10 20 16 18 24 33 47 74
FALSE: 76 83 45 90 66 33 72 84

From above list, $X1$=10 and $X2$=90. The calculation of $S(x_i)$ is shown in Table 1.

The minimum occurs at x_2, i.e. at 26. Thus. 36 can be used as a good threshold value. This result can be compared with the median. Now, 36 is a border line for 1 and 0 logic, so

Table 1. Calculation of the Threshold Value

x_i	x_1	x_2	x_3	x_4	x_5
Value for x_i	8	26	40	60	75
Range for $p_k(x_i)$	10 - 18	10 - 36	10 - 50	10 - 70	10 - 85
$p_T(x_i)$	3/3	6/7	7/9	7/10	8/15
$p_F(x_i)$	0/3	1/7	2/9	3/10	7/15
$q_T(x_i)$	5/13	2/9	1/7	1/6	0/1
$q_F(x_i)$	8/13	7/9	6/7	5/6	1/1
$p(x_i)$	3/16	7/16	9/16	10/16	15/16
$q(x_i)$	13/16	9/16	7/16	6/16	1/16
$S_p(x_i)$	0.000	0.400	0.529	0.611	0.691
$S_q(x_i)$	0.666	0.530	0.410	0.451	0.000
$S(x_i)$	0.541	0.473	0.476	0.551	0.648
Threshold		x			

the 1/0 table is:
True: 0 0 0 0 0 0 1 1
False: 1 1 1 1 1 0 1

Assuming that only this digit is used for separation, the rule is:

0 → True, or 1 → False.

We have two wrong decisions in True and one wrong decision in False. But if we choose the border line the median, i.e.. 47 in this example, we have two wrong decisions in True and another two wrong decisions in False. The 1/0 table is omitted for the case of median. Here it is shown that the threshold value is better than any arbitrarily chosen value in this simple example. But the more important point here is that we have a basis for finding a border line of 1 and 0 that leads us into a better position in rule induction. As $S(x_3)$ indicates, a border line value of 50 as another threshold value has the very same result as the value of 36.

Thus the induced rule will be simpler and more reliable with this threshold value.

DETECTION AND LEARNING

Induction with entropy minimization is basically for pattern recognition, that is to classify faults and switching events. However, we can apply this tool for detection to the learning problem. Information-processing systems that improve their performance or enlarge their knowledge bases are said to be "learning"[7]. One of the objectives of learning is to automate the acquisition of knowledge. The relationship of detection and learning with minimum entropy is summarized as below:

Detection:

1) Get as many samples as possible on faults and switching events.
2) Get as many independent variables as possible to be used to separate samples into two classes
3) List all the samples with their corresponding values for each independent variable
4) Find a threshold value for each independent variable with minimum entropy
5) Find 1/0 table with the threshold value given above
6) Perform the procedure of finding a rule with minimum entropy
7) Use only this rule as a detection tool.

All the steps except 7) are preliminary and preparatory states. Only the rule is used to detect faults and switching events.

Learning:

1) Write a learning program which reads variable values of sample. The variables are given by a human. The class of the sample is also given by a human expert.
2) At the end of putting samples into program, the learning program will do jobs 3) through 6) in the detection procedure.
3) Finally the learning program has a rule which can classify samples.
4) Use this rule for detection.

Figure 2 shows the relationship between detection and learning. Stage 1 is for learning and stage 2 is for detection. At stage 1, from parameters, a human expert chooses the variables for detection. Whenever the sample event comes into the learning program with the variables chosen, an expert teaches the learning program which sample event is "Event" and which is "Fault". If all of the sample events are entered, the learning program performs sample event listing–Threshold value calculation–1/0 table–Rule induction with minimum entropy. So, the output of the learning program is the induced rule. Additionally, from the induced rule we can see which variables are essential in detection and which are not. At stage 2, the induced rule from stage 1 is used. The same variables chosen by an expert can be used, or only variables which proved to be essential in detection from the induced rule at stage 1 can be used.

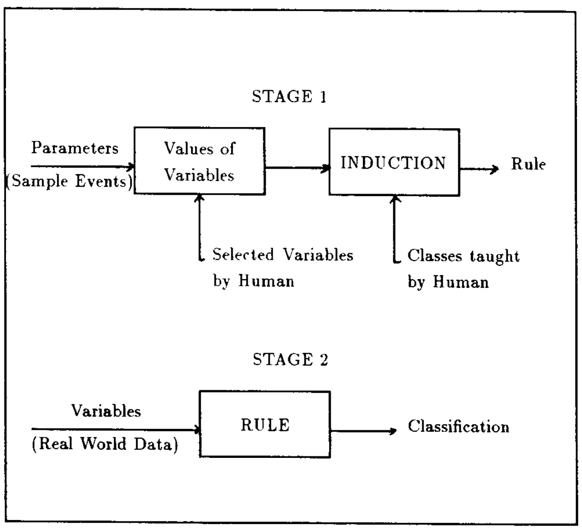


Figure 2. Learning and Detection

APPLICATION EXAMPLE

Twenty-one faults on various unusual surface conditions have been studied. The purpose is not to discriminate faults from normal activity, but to discriminate faults on unusual surface conditions(i.e., grass, asphalt, concrete) from switching events.

The surface conditions involved were reinforced concrete, non-reinforced concrete, dry asphalt, wet macadam, and grass. For switching events, 10 samples of various switching conditions were studied. Switching conditions included are air switch on and off, capacitor bank on and off, and load tap changer raise and lower. We focus on the frequency characteristics of the fault and switching event waveforms. For the frequency components, all of the harmonics can be examined[8]. But here we chose only odd harmonics. The measurement parameter used for the odd harmonics is the "energy" level.

The independent variables we chose are described below. Each independent variable occupies one digit.

Digit 1: The number of burst cycles(on-cycles) in 30 cycle window. Here, it should be mentioned that a long term average energy of normal states is known and if the energy is 1.5 times greater that that of normal, it is counted as an "on" cycle; otherwise, it is counted as an "off" cycle.
Digit 2: The number of off-cycles in the same window length.
Digit 3: The average number of on-cycles.
Digit 4: The average number of off-cycles.
Digit 5: Randomness index.
Digit 6: Average relative amplitude increase of on-cycles.
Digit 7: Average relative amplitude increase of off-cycles.

Digits 1, 2, 3, and 4 show both their activities and sporadicities. Digits 6 and 7 are essential to detect level changes of the switching events. Digit 5 is defined here: If the energy of a current cycle deviates from the energy of the previous cycle by 0.5 times the normal energy level, count it as an index of randomness.

All of the procedures which find the values of 7 digits start when the the first on-cycle is seen. The list of all 31 samples is shown in Table 2. We have 21 fault cases on various surface conditions such as concrete, asphalt, and grass. Ten switching event cases include air switch operations, capacitor bank operations, and a load tap change.

There may be many ways to set steps(x_i) between $X1$ and $X2$. For simplicity, we have performed the calculation of each threshold value with the following steps.

We have 10 steps for each independent variable to find a threshold value. $X1$ is minimum value of each independent variable and $X2$ is the maximum of it. The single step value is derived by $(X2 - X1)/10$. The procedure of threshold value calculation for the first digit is shown at Table 3. The samples of Event and Fault are shown below:

EVENT: 3 4 4 2 30 30 30 2 1 8
FAULT: 9 8 6 20 13 2 1 15 27 2 1 1 2 1 2 30 2 3 3 2 2

$S(x_9)$ has a minimum entropy. Therefore the threshold value is 28.

Other threshold values are similarly calculated and each value of step x_i and corresponding entropy from digits 2 through 7 is shown at Table 4. The values in bold face indicate the minimum entropies.

Now we have threshold values for each digit. If any value is greater than its threshold value at each digit, that value will be converted to 1 in the 1/0 table, otherwise 0.

Below is the list of the threshold values:

Digits:	1	2	3	4	5	6	7
Threshold Value:	28	9	22	27	2	1.79	1.01

Table 2. The List of Samples

#	DIGIT 1	2	3	4	5	6	7	Class
1	9	21	2.25	7	7	1.74	1.01	Fault
2	8	22	1.60	4.4	3	1.82	1.04	Fault
3	6	24	1.50	8	4	1.86	1.07	Fault
4	20	10	20.0	10	0	1.67	1.35	Fault
5	13	17	4.30	8.5	0	1.58	1.35	Fault
6	2	28	2	28	0	1.60	1.19	Fault
7	1	29	1	29	0	1.53	1.06	Fault
8	15	15	7.5	15	0	1.78	1.09	Fault
9	27	3	13.5	3	1	2.08	1.45	Fault
10	2	28	2	28	1	1.58	1.04	Fault
11	1	29	1	29	1	1.61	0.99	Fault
12	1	29	1	29	0	1.55	0.97	Fault
13	2	28	2	28	1	1.56	1.01	Fault
14	1	29	1	29	0	1.58	1.05	Fault
15	2	28	2	28	1	1.61	1.04	Fault
16	30	0	30	0	0	2.14	1.00	Fault
17	2	28	2	28	0	1.62	1.06	Fault
18	3	27	3	27	1	1.66	1.05	Fault
19	3	27	3	27	1	1.62	1.07	Fault
20	2	28	2	28	2	1.73	1.05	Fault
21	2	28	2	28	2	1.99	0.99	Fault
22	3	27	3	27	4	2.08	0.99	Event
23	4	26	4	26	5	2.59	1.14	Event
24	4	26	4	26	3	2.07	0.89	Event
25	2	28	1	14	4	2.22	0.88	Event
26	30	0	30	0	9	10.23	1.00	Event
27	30	0	30	0	8	10.43	1.00	Event
28	30	0	30	0	8	10.34	1.00	Event
29	2	28	2	28	4	1.91	0.20	Event
30	1	29	1	29	4	0.79	0.17	Event
31	8	22	1.3	3.7	0	1.52	1.43	Event

Table 3. Threshold Value Calculation for First Digit

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Value of x_i	3	6	9	12	15	18	21	24	27
Range	1 - 4	1 - 7	1 - 10	1 - 13	1 - 16	1 - 19	1 - 22	1 - 25	1 - 28
$p_E(x_i)$	6/19	6/20	7/23	7/24	7/25	7/25	7/26	7/26	7/27
$p_F(x_i)$	13/19	14/20	16/23	17/24	18/25	18/25	19/26	19/26	20/27
$q_E(x_i)$	4/12	4/11	3/8	3/7	3/6	3/6	3/5	3/5	3/4
$q_F(x_i)$	8/12	7/11	5/8	4/7	3/6	3/6	2/5	2/5	1/4
$p(x_i)$	19/31	20/31	23/31	24/31	25/31	25/31	26/31	26/31	27/31
$q(x_i)$	12/31	11/31	8/31	7/31	6/31	6/31	5/31	5/31	4/31
$S_p(x_i)$	0.624	0.611	0.614	0.603	0.593	0.593	0.582	0.582	0.572
$S_q(x_i)$	0.637	0.656	0.662	0.683	0.694	0.694	0.673	0.673	0.563
$S(x_i)$	0.630	0.627	0.626	0.621	0.613	0.613	0.597	0.597	0.571
Threshold									x

The 1/0 table with the calculated threshold values is shown in Table 5.

The steps of rule derivations are shown below. The reliability of the rule is also shown.

First. the digit count, d_n, needs to be found for Event and Fault classes.

For Event:$d_1 = 0.30, d_2 = 0.70, d_3 = 0.30, d_4 = 0.20, d_5 = 0.90, d_6 = 0.80, d_7 = 0.20$

For Fault:$d_1 = 0.95, d_2 = 0.10, d_3 = 0.95, d_4 = 0.52, d_5 = 0.86, d_6 = 0.81, d_9 = 0.29$

Then, the added d_n's are shown:

$d_1 = 1.25, d_2 = 0.80, d_3 = 1.25, d_4 = 0.72, d_5 = 1.76, d_6 = 1.61, d_7 = 0.49$

Table 4. Calculation of Threshold Value for the Other Digits

	DIGITS 2	3	4	5	6	7
x_1	3	3	3	1	1	0.12
$S(x_1)$	0.618	0.626	0.597	0.417	**0.482**	0.551
x_2	6	6	6	2	2	0.24
$S(x_2)$	**0.597**	0.621	0.613	**0.344**	0.509	0.551
x_3	9	9	9	3	3	0.36
$S(x_3)$	**0.597**	0.613	0.624	0.379	0.509	0.551
x_4	12	12	12	4	4	0.48
$S(x_4)$	0.613	0.613	0.616	0.534	0.509	0.551
x_5	15	15	15	5	5	0.60
$S(x_5)$	0.621	0.600	0.616	0.571	0.509	0.551
x_6	18	18	18	6	6	0.72
$S(x_6)$	0.621	0.600	0.616	0.509	0.509	0.561
x_7	21	21	21	7	7	0.84
$S(x_7)$	0.628	**0.571**	0.616	0.509	0.509	**0.507**
x_8	24	24	24	8	8	0.96
$S(x_8)$	0.629	**0.571**	0.616	0.591	0.509	0.629
x_9	27	27	27	9	9	1.08
$S(x_9)$	0.655	**0.571**	**0.579**	0.628	0.509	0.621

Table 5. The 1/0 Table

#	DIGITS 1	2	3	4	5	6	7	class	#	DIGITS 1	2	3	4	5	6	7	class
1	0	1	0	0	1	0	0	F	17	0	1	0	1	0	0	1	F
2	0	1	0	0	1	1	1	F	18	0	1	0	0	0	0	1	F
3	0	1	0	0	1	1	1	F	19	0	0	0	0	0	1	1	F
4	0	1	0	0	0	0	1	F	20	0	1	0	1	0	0	1	F
5	0	1	0	0	0	0	1	F	21	0	1	0	0	0	1	0	F
6	0	1	0	1	0	0	1	F	22	0	1	0	0	1	1	0	E
7	0	1	0	1	0	0	1	F	23	0	1	0	0	1	1	1	E
8	0	1	0	0	0	0	1	F	24	0	1	0	0	1	1	0	E
9	0	0	0	0	0	1	1	F	25	0	1	0	0	1	1	0	E
10	0	1	0	1	0	0	1	F	26	1	0	1	0	1	1	0	E
11	0	1	0	1	0	0	0	F	27	1	0	1	0	1	1	0	E
12	0	1	0	1	0	1	0	F	28	1	0	1	0	1	1	0	E
13	0	1	0	1	0	1	0	F	29	0	1	0	1	1	1	0	E
14	0	1	0	1	0	0	1	F	30	0	1	0	1	1	0	0	E
15	0	1	0	1	0	0	1	F	31	0	1	0	0	0	0	1	E
16	1	0	1	0	0	1	0	F									

Then the indicies of index are calculated and shown:

$I_1 = 0.25, I_2 = 0.20, I_3 = 0.25, I_4 = 0.28, I_5 = 0.76, I_6 = 0.61, I_7 = 0.51$

The fifth digit has the maximum index of all digits. The first step is to use this digit. Before we perform the steps, we want to define some variables here. x is the number of samples in the digit formation for each step in either Event or Fault. n is the number of samples in the digit formation for each step in both Event and Fault.

xxxx1xx: Both(Event:$x/n = 9/12$)
xxxx0xx: Both(Fault:$x/n = 18/19$)
Therefore, the first step is : xxxx0xx $\rightarrow$ Fault
Then the mean probability of this step is:
$< p >= (x+1)/(n+2) = (18+1)/(19+2) = 0.90$.

We eliminate all the samples of this digit formation, then we have following remainders.

Event: 0100110, 0100111, 0100110, 0100110, 1010110
1010110, 1010110, 0101110, 0101100
Fault: 0100100, 0100111, 0100111

Similarly the indicies of index are calculated, and shown:
$I_1 = 0.33, I_2 = 0.33, I_3 = 0.33, I_4 = 0.22, I_5 = 0.00, I_6 = 0.22, I_7 = 0.56$

Thus, the 7th digit has the maximum index. So
xxxxxx1: Both(Fault:$x/n = 2/3$)
xxxxxx0: Both(Event:$x/n = 8/9$)
So the second step is : xxxxxx0 $\rightarrow$ Event
Then $<p>= 9/11 = 0.82$

Then, remainders are:
Event: 0100111
Fault: 0100111, 0100111

So the third step is simple
xxxxxxx $\rightarrow$ Fault($x/n = 2/3$) with $<p>= 0.6$

The induced rule is shown below:
Step1: xxxx0xx $\rightarrow$ Fault, $<p>= 0.90, x = 18$
Step2: xxxxxx0 $\rightarrow$ Event, $<p>= 0.82, x = 8$
Step3: xxxxxxx $\rightarrow$ Fault, $<p>= 0.60, x = 2$

Thus, the entropy of this rule is,
$S = -k((18)0.9\ln 0.9 + (8)0.82\ln 0.82 + (2)0.6\ln 0.6) = 3.62k$

The specific entropy is defined as[3] S/W, where W is the number of samples, is $S/W = 3.62k/28 = 0.13k$

The reliability of the rule is also defined as[3];
$-k\ R\ \ln R = S/W$,
so reliability of this rule is $R = 0.86$.

Applying this rule to detection, the first step is to find randomness. If it is 0 in the 1/0 table, it is considered as Fault. If not 0, then move to second step to check whether the average amplitude increase of off-cycles is 0 in the 1/0 table. If it is, then that is Event. If not, the third and final step is to consider all the remainders as Fault. On the steps of the induced rule, it is seen that digits 1 through 4 and digit 6 are not important in discrimination of high impedance faults on unusual surface conditions from switching events. The most important digit is the randomness index.

From this rule, we can detect and classify samples. The other aspect of this rule is that it determines which variables are important and which are not for classification purposes.

CONCLUSIONS

As a detection/classification tool for various fault data, the induction rule with minimum entropy is proposed. This method is easy and very promising, especially when it is hard to set probability factors or deterministic decision criteria on the sample data, and holds promise in the classification of high impedance faults. This induction rule for pattern recognition does not need any probability threshold. The learning ability to improve the performance of detection is briefly mentioned. For a better induction rule, setting proper threshold values for the continuous world is also proposed. This detection tool is proposed primarily for comparing faults on unusual surface conditions with switching events, but this tool can also be to compare and classify all kinds of high impedance faults and switching events as well as normal states. An area for additional study is the setting of values for defining on-cycles and off-cycles which affect the detection process.

The methods presented in this paper are taken from ongoing research in high impedance fault detection. While the techniques have not been reduced to practice or field tested, they hold promise for future improvements in the relaying of high impedance faults.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support of the National Science Foundation, the Electric Power Research Institute, and the Center for Energy and Mineral Resources. Without their support, the research investigations of the Power System Automation Laboratory of the Electric Power Institute would not be possible.

REFERENCES

[1]. B. G. Buchanan, E. H, Shortliffe(Ed.), *Rule-Based Expert Systems*, Addison-Wesley, 1985.

[2]. B. Don Russell, Ram P. Chinchali, C. J. Kim, "Behavior of Low Frequency Spectra during Arcing Fault and Switching Event," *IEEE* PES Summer meeting, San Francisco, CA, 1987. Paper No. 87SM633-1.

[3]. Ronald Christensen, *Entropy Minimax Sourcebook*, Entropy Limited, 1980.

[4]. Maximum Entropy Conference:held at MIT, MIT Press, 1978

[5]. B. Don Russell, K. Mehta, Ram P. Chinchali, "An Arcing Fault Detection Technique using Low Frequency Current Components–Performance Evaluation using Recorded Field Data", *IEEE* PES Summer Meeting, San Francisco, CA, 1987. Paper No. 87SM634-9.

[6]. C. L. Benner, P. W. Carswell, B. Don Russell, "Improved Algorithm for Detecting Arcing Faults using Random Fault Behavior", Submitted to *Electric Power System Research.*

[7]. S. L. Tanimoto, *The Elements of Artificial Intelligence*, Computer Science Press, 1987.

[8]. C. J. Kim, B. Don Russell, "Harmonic Behavior during Arcing Faults on Power Distribution Feeders", *Electric Power System Research*, Vol. 14, No.3, pp. 219-225, June 1988.

BIOGRAPHY

C. J. Kim, received his BSEE and MSEE degrees from Seoul National University, Seoul, Korea in 1980 and 1982, respectively. He is currently working toward his Ph.D. degree at Texas A&M University. During 1983-1985, he worked on the computerized power system monitoring for Gold Star Instrument Research Center, Korea. Mr. Kim's research interests are AI and expert system applications in power system monitoring, protection, and control.

B. Don Russell(SM), received BS and ME degrees in Electrical Engineering at Texas A&M University. He holds a Ph.D. from University of Oklahoma in power system engineering. Dr. Russell is Professor of Electrical Engineering and directs the activities of Power System Automation Laboratory of the Electric Power Institute, Texas A&M University. His research centers on the use of advanced technologies to solve problems in power system control, protection, and monitoring. He is the recipient of several awards for advanced technology applications. Dr. Russell chairs several working working groups and subcommittees of PES. Dr. Russell is a member of the Substation Committee and Power Engineering Education Committee of PES. He is a registered professional engineer and a director of the Texas Society of Professional Engineers.