Comparison of Bayesian and Fuzzy ARTmap Networks in HV Transmission Lines Fault Diagnosis

HAMED NAFISI[#1], BABAK ABDI[#1], AMIRHOSSEIN AGHAKHANI[#2]
[#1]Department of Electrical Engineering
Islamic Azad University-Damavand Branch
Near Moallem Square, Damavand
IRAN
[#2]Department of Electrical Engineering
Amirkabir University of Technology
No. 424, Hafez Avenue, Tehran
IRAN
nafisi@aut.ac.ir babakabdii@gmail.com a.h.aghakhani@aut.ac.ir

Abstract: - Fault diagnosis is a vital discussion in power systems restoration. Recently, much research endeavors have been done for fault section diagnosis of power systems by using several techniques, such as rule-based expert system, logic-based expert system, fuzzy relation based expert system, neural network, optimization techniques based approach, etc. They diagnose the fault from different ways. However, each approach has its limitations. In this paper, a Bayesian approach by RBF learning using a simulation technique, the Markov chain Monte Carlo (MCMC) and Fuzzy ARTmap network are proposed to predict the fault in a typical power transmission line and the results are compared.

Key-Words: - Bayesian Network, Fuzzy ARTmap, Transmission Line, Fault Diagnosis, MCMC, Variable Thevenin

1 Introduction

HV Transmission lines are one of the most essential equipments in the electrical power systems. While a short circuit in a power system is occurred, protective relays will operate and clear the fault. Keeping in mind that in larger power systems, multi circuits' transmission lines, more relays, registers, circuit breakers, etc are exist. So, number of warnings dispatched to control center during the fault accruing is much more. In case those main protections do not operate properly and on-time, back up protections will operate related to this situation. There are two main negative consequences: first, it causes undesirable load interruptions of other buses, and second, leads to decision making of operator in emergency conditions to be more difficult, according to variety of dispatched warnings to main control center.

Due to it involving a lot of uncertain signals, which are caused by many factors, such as mal-operation and non-operation of circuit breakers or relays, data-transmission error of loss, the inaccurate time of the protective operation, fault diagnosis of network needs uncertainty reasoning. Among the existing uncertainty reasoning approaches, Bayesian networks approach stands out as the only one that is directly grounded in probability theory. Bayesian network based approach, mainly used for representing and reasoning with uncertainty, has been successfully used in many fields, such as speech recognition, industrial control, economic forecasting, as well as medical diagnosis [1]. Recently, with the development of Data Mining, the capabilities of inference and learning of Bayesian network have gained more and more attention. However, little researches have been done to address how to use Bayesian network in power systems.

Moreover, one of good neural networks which is not used yet in power systems is Fuzzy ARTmap which is a developed member of ART neural networks family.

This paper will focus on the construction and application of Bayesian network and Fuzzy ARTmap models. The Bayesian network (e.g., [2] and [3]) is a probabilistic graphical model in which a problem is structured as a set of variables (parameters) and probabilistic relationships among them. The constructed Bayesian network, after serial evidence–propagation inferences, sorts the transmission line faults according to their individual some more important harmonics occurrence probability. The major contributions of this paper are as follows.

In this paper, a Bayesian network and Markov chain Monte Carlo (MCMC) simulation technique is compared to Fuzzy ARTmap network in order to predict the faulted phase or phases in a typical high voltage transmission line.

The format of this paper is as follows. Section 2 and 3 present the fundamental of Bayesian and Fuzzy ARTmap networks, respectively. In Section 2, the proposed model structuring procedure and data assessment based on Radial Basis Function (RBF) method [4] will be presented. In the section 4, the transmission line's model and data required for Bayesian and FAM networks will be presented. The test results of constructed Bayesian and Fuzzy ARTmap network will be followed and concluding remarks will be given.

2 Fundamental of Bayesian Network

Bayesian networks are usually used to model the situations (e.g., medical diagnosis) in which causality plays a role but where the understanding of what is actually going on is incomplete. That is, a Bayesian network for the domain represents a joint probability distribution over a set of variables (i.e., chance nodes).

Bayesian network is a directed acyclic graph that consists of single-evidence, multiple-evidence, and multiple-layer probabilistic relationships among the variables. For detailed description about DAGs see [5]. Thus, Bayesian network expresses the global joint distribution with a set of local distributions and relates only the neighboring nodes. Fig. 1 illustrates the basic structure of a Bayesian network.

The Bayesian network has been successfully applied in many fields such as medical diagnosis [6], equipment diagnosis [7], and mineral exploration [8]. Extensive review of Bayesian networks can be found, for example, in [9].

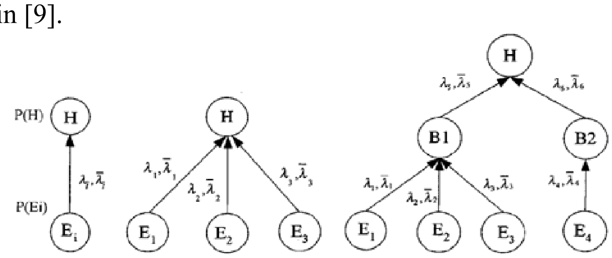

Fig.1 Basic structures of Bayesian network (DAG) [5]

The Bayesian network is a directed acyclic graph in which the following holds:

- A set of random variables makes up the nodes of the network.
- A set of directed links or arrows connects pairs of nodes.
- Each node has a conditional probability table that quantifies the effects that the parents have on the node. The parents of a node are all those nodes that have arrows pointing to it.
- The graph has no directed cycles (hence is a directed, acyclic graph or DAG).

A Bayesian network provides a complete description of the domain. Every entry in the joint probability distribution can be calculated from the information in the network. A generic entry in the joint is the probability of a conjunction of particular assignments to each variable. The value of this entry is given by the following formula [10]:

$$P(x_1,...,x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i)) \qquad (1)$$

We use the notation $P(x_1,...,x_n)$ as an abbreviation for this. Thus, each entry in the joint is represented by the product of the appropriate elements of the conditional probability tables (CPTs) in the belief network. The CPTs therefore provide a decomposed representation of the joint.

$$\begin{aligned} &P(x_1,...,x_n) = \\ &P(x_n \mid x_{n-1},...,x_1)P(x_{n-1} \mid x_{n-2},...,x_1)...P(x_1) = \\ &\prod_{i=1}^{n} P(x_i \mid x_{i-1},...,x_1) \end{aligned} \qquad (2)$$

Then we repeat this process, reducing each conjunctive probability to a conditional probability and a smaller conjunction. We end up with one big product. Comparing this with Equation (1) and (2), we see that the specification of the joint is equivalent to the general assertion that.

2.1 Markov Chain Monte Carlo blanket

A node is conditionally independent of its non-descendants, given its parents.

A node is conditionally independent of all other nodes in the network, given its parents, children, and children's parents that is, given its Markov blanket.

From these conditional independence assertions and the CPTs, the full joint distribution can be reconstructed; thus, the "numerical" semantics and the "topological" semantics are equivalent.

According to the theory of Markov blanket, the nodes for inference, such as the fault node or protection node, are chosen first.

2.2 The MCMC Algorithm

The MCMC generates each event by making a random change to the preceding event. It is therefore helpful to think of the network as being in a particular current state specifying a value for every variable. The next state is generated by randomly sampling a value for one of the non-evidence variables X_i, conditioned on the current values of the variables in the Markov blanket of X_i. MCMC therefore wanders randomly around the state space-the space of the possible complete assignments-flipping one variable at a time, but keeping the evidence variables fixed. The algorithm is that:

Let $q(x \to x')$ be the probability that the process makes a transition from states x to state x'. This transition probability defines what is called a Markov chain on the state space. Now suppose that we run the Markov chain for t steps, and let $P_t(x)$ be the probability of being in state x at time t. Similarly, let $P_{t+1}(x')$ be the probability of being in state x' at time $t+1$. Given $P_t(x)$, we can calculate $P_{t+1}(x')$ by summing, for all states the system could be in at time t, the probability of being in that state times the probability of making the transition to x':

$$P_{t+1}(x') = \sum_x P_t(x) q(x' \to x) \quad (3)$$

We will say that the chain has reached its stationary distribution if $P_t(x) = P_{t+1}(x')$. Let us call this stationary distribution P; its defining equation is therefore:

$$P(x') = \sum_x P(x) q(x' \to x) \quad (4)$$

For all x'

Under certain standard assumptions about the transition probability distribution q, there is exactly one distribution P satisfying this equation for any given q. Equation (3) can be read as saying that the expected "outflow" from each state (i.e., its current "population") is equal to the expected "inflow" from all the states. One obvious way to satisfy this relationship is if the expected flow between any pair of states is the same in both directions. This is the property of detailed balance:

$$P(x') q(x \to x') = \sum_x P(x) q(x' \to x) \quad (5)$$

For all x, x'

2.3 Radial Basis Function (RBF)

In order to assess the input data Radial Basis Function (RBF) method is used. A radial basis function (RBF) is a real-valued function whose value depends only on the distance from the origin, so that [4]:

$$\phi(x) = \phi(\| x \|) \quad (6)$$

Or alternatively on the distance from some other point *c*, called a *center*, so that

$$\phi(x, c) = \phi(\| x - c \|) \quad (7)$$

Any function ϕ that satisfies the property $\phi(x) = \phi(\| x \|)$ is a radial function. The norm is usually *Euclidean distance*.

Radial basis functions are typically used to build up function approximations of the form:

$$y(x) = \sum_{i=1}^{N} \omega_i \phi(\| x - c_i \|) \quad (8)$$

Where the approximating function $y(x)$ is represented as a sum of N radial basis functions, each associated with a different center c_i, and weighted by an appropriate coefficient ω_i. Approximation schemes of this kind have been particularly used in time series prediction and control of nonlinear systems exhibiting sufficiently simple chaotic behaviour.

The sum can also be interpreted as a rather simple single-layer type of artificial neural network called a radial basis function network, with the radial basis functions taking on the role of the activation functions of the network. It can be shown that any continuous function on a compact interval can in principle be interpolated with arbitrary accuracy by a sum of this form, if a sufficiently large number N of radial basis functions is used.

There are some commonly used types of radial basis functions include $r = \| x - c_i \|$ which as follows:

- Gaussian:
  $$\phi(r) = \exp(-\beta r^2) \quad \text{for some } \beta > 0 \quad (9)$$
- Multi-quadric:
  $$\phi(r) = \sqrt{r^2 + \beta^2} \quad \text{for some } \beta > 0 \quad (10)$$
- Polyharmonic spline:
  $$\phi(r) = r^k \quad k=1, 3, 5, \ldots$$
  $$\phi(r) = r^k \ln(r) \quad k=2, 4, 6, \ldots \quad (11)$$
- Thin plate spline (a special polyharmonic spline):
  $$\phi(r) = r^2 \ln(r) \quad (12)$$

These polyharmonic splines (which include the thin-plate spline) minimise certain energy semi-norms and are therefore the "smoothest" interpolators. Note that the associated basic functions are not compactly supported - *they grow as r increases from the origin*.

RBFs are popular for interpolating scattered data as the associated system of linear equations is guaranteed to be invertible under very mild conditions on the locations of the data points. For example, the thin-plate spline only requires that the points are not co-linear while the Gaussian and multi-quadric place no restrictions on the locations of the points. In particular, RBFs do not require that the data lie on any sort of regular grid.

In this paper Radial Basis Function method with Gaussian type is considered for assessment of data.

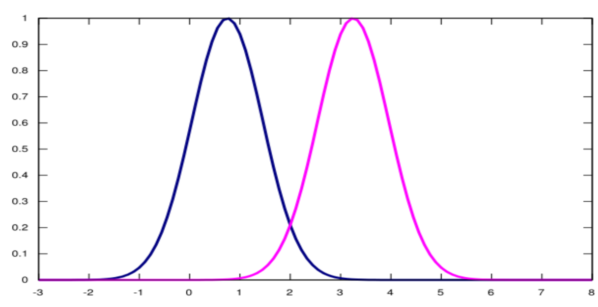

Fig.2 Un-normalized Radial Basis Functions with c_1= 0.75 and c_2=3.25

2.3.1 Estimating the weights

The approximant $y(x)$ is differentiable with respect to the weights ω_i. The weights could thus be learned using any of the standard iterative methods for neural networks. But such iterative schemes are not in fact necessary: because the approximating function is *linear* in the weights ω_i, the ω_i can simply be estimated directly, using the matrix methods of linear least squares.

The input assessment has been shown in this section. In the next section required input-output data and modeling of power system will be presented.

3 Fuzzy ARTmap Neural Network

For the purpose of training and testing, in this paper Adaptive Resonance Theory (ART) neural networks have been used. In general, this family of neural networks include ART1, ART2 [14], ART3 [15], ARTmap [16], Fuzzy ART [17] and Fuzzy ARTmap [18]. ART1 and ARTmap categorize the binary input patterns while, Fuzzy ARTmap are also capable to categorize analogue patterns.

Fuzzy ARTmap is an incremental supervised learning algorithm which combines fuzzy logic and Adaptive Resonance Theory (ART) neural network for recognition of pattern categories and multidimensional maps in response to input vectors presented in an arbitrary order. It realizes a new minmax learning rule which conjointly minimizes predictive error and maximizes code compression, and therefore gives generalization. This is achieved by a match tracking process that increase the ART vigilance parameter (fuzzy degree of membership of the input with respect to the category templates) by the minimum amount needed to correct a predictive error (PE). The Fuzzy ARTmap neural network is composed of two Fuzzy ART modules [18], i.e. *fuzzy* ART_a and *fuzzy* ART_b, which are depicted in Fig.3 and are essentially the same as those described by Carpenter et al.

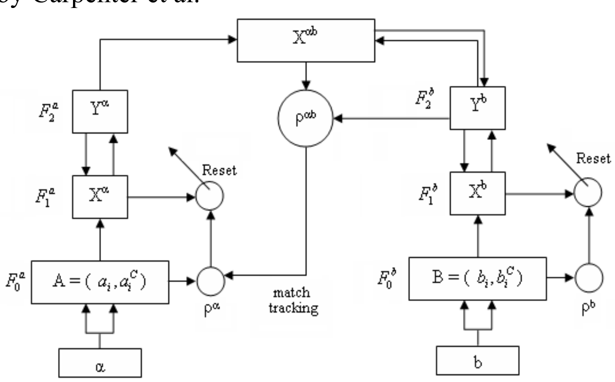

Fig.3 Typical Fuzzy ARTmap

The interactions mediated by the map field F^{ab} operationally characterized as follows.

3.1 ART_a and ART_b

Inputs to ART_a and ART_b are in the complement code form: for ART_a ‹$I = A = (a,a^c)$ and for ART_b ‹$I = B = (b,b^c)$ (See Fig.3). Variables in ART_a or ART_b are designated by subscript "a" and "b" respectively. For ART_a, let $x^a = \{x_1^a,\ldots,x_{2Ma}^a\}$ denote the F_1^a output vector, let $y^a = \{y_1^a,\ldots,y_{Na}^a\}$ denote F_2^a, and let $w_j^a = \{w_{j1}^a,\ldots,w_{j2Ma}^a\}$ denote the j^{th} ART_a weight vector. For ART_b, let $x^b = \{x_1^b,\ldots,x_{2Mb}^b\}$ denote the F_1^b output vector and let $y^b = \{y_1^b,\ldots,y_{Nb}^b\}$ denote F_2^b. And let $w_k^b = \{w_{k1}^b,\ldots,w_{k2Mb}^b\}$ denote the k^{th} ART_b weight vector. For the map field, let $x^{ab} = \{x_1^{ab},\ldots,x_{Na}^{ab}\}$ denote the F^{ab} output vector, and let $w_j^{ab} = \{w_{j1}^{ab},\ldots,w_{jNb}^{ab}\}$ denote the weight vector from the j^{th} F_2^a node to F^{ab}. Vectors x^a, y^a, x^b, y^b, and x^{ab} are set to 0 between input presentations.

3.2 Map field activation

The map field F^{ab} is activated whenever one of the ART_a or ART_b categories is active. If node J of F_2^a is chosen, then its weights w_j^{ab} activate F^{ab}. If K in F_2^b is active, then node K in F^{ab} is activated by one-to-one pathways between F_2^b and F^{ab}. If both ART_a and ART_b are active, then F^{ab} becomes active only if ART_a predicts the same category as ART_b via the weights w_j^{ab}. The F^{ab} output vector x^{ab} obeys the following:

$$\begin{cases} y^b \wedge w_j^{ab} & \textit{If the } j^{th} F_2^a \textit{ is active and } F_2^b \textit{ is active.} \\ w_j^{ab} & \textit{If the } j^{th} F_2^a \textit{ is active and } F_2^b \textit{ is inactive.} \\ y^b & \textit{If the } j^{th} F_2^a \textit{ is inactive and } F_2^b \textit{ is active.} \\ 0 & \textit{If the } j^{th} F_2^a \textit{ is inactive and } F_2^b \textit{ is inactive.} \end{cases} \quad (13)$$

From (13), $x^{ab} = 0$ if the prediction w_j^{ab} is disconfirmed by y^b. Even such a mismatch triggers and ART_a search for a better category, as follows.

3.3 Match tracking

At the start of each input presentation, the ART_a vigilance parameter ρ_a equals to baseline vigilance ρ_a. The map field vigilance parameter is ρ_{ab}

$$If \quad |x^{ab}| < \rho_{ab} \cdot |y^b| \quad (14)$$

The ρ_a is increased until it is slightly larger than $|A \wedge w_j^a||A|^{-1}$, where A is the input to F_1^a, in complement coding form, and

$$|x^a| = |A \wedge w_j^a| < \rho_a |A| \quad (15)$$

Where J is the index of active F_2^a node.

When this occurs, ART_a search leads either to activation of another F_2^a node J with:

$$|x^a| = |A \wedge w_j^a| \geq \rho_a |A| \quad (16)$$

and

$$\left|x^{a}\right| = \left|y^{b} \wedge w_{j}^{ab}\right| \geq \rho_{a}\left|y^{b}\right| \tag{17}$$

Or, if no such nodes exist, i.e. the input pattern to F_2^a layer does not match any pattern: the input pattern is classified as a new pattern.

4 Model Structure for Power System

Considered power transmission line in this paper is a 400KV, single circuit, 4 bundled and transposed, with 160km length. Modelling of each part of this system shall be done according to the following.

4.1 Transmission Line Modelling

In order to simulate the mentioned power system, EMTP software is used to analyse needed data. In this software there are some transmission line models such as: distributed models, Meyer-Dommel, Semlyen, and J-Marti model. Because of accuracy of J-Marti, this model is used in this paper for modelling of transmission line.

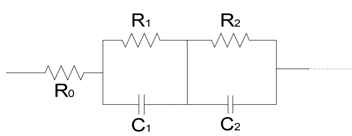

Fig.4 Equivalent circuit of characteristic impedance of J-Marti model

4.2 Variable Thevenin model for Busbars

Although the transmission line in the nominal power system's frequency has an inductive behaviour, but in higher frequencies it may has capacitive behaviour in some frequencies. Moreover, in some frequencies it has pure resistive impedance [11]. Not only simulation of all parts of a three phase AC power system is easy but also it is not economical, therefore it is better to use RLC parallel branches with same responses in a wide range of frequencies for simulation of behaviour of the system. Impedance-Frequency curve of variable thevenin model is depicted in Fig.5.

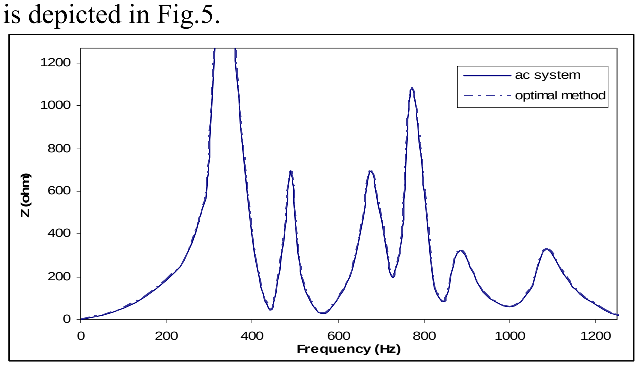

Fig.5 Impedance-Frequency curve of variable thevenin model

Other required data for transmission line and variable thevenin model of busbars is given in [11].

In this paper Fast Fourier Transform (FFT) technique is used for making harmonics of voltages and currents of the sending bus for a period (20 msec.) post of fault time [12]. In this way, first to ninth harmonics plus DC component of waveforms of voltages and currents are used for inputs of neural networks. These values are earned for 10 exists type of faults; 3 single phase fault, 3 double phase fault, 3 double phase to ground fault and a three phase fault. Moreover than this, these faults are simulated for 0, 45, and 90 degree of phase for each 10 km fault of line. So, there is 410 data for train and near to 100 cases for test of the network. The outputs of the network are divided to 11 classes of fault diagnosis; 10 classes of faults and a class for normal conditions.

A sample of waveforms of a fault is illustrated in the following figures. As it is clear in the figures these faults made harmonics that those are useful for studies.

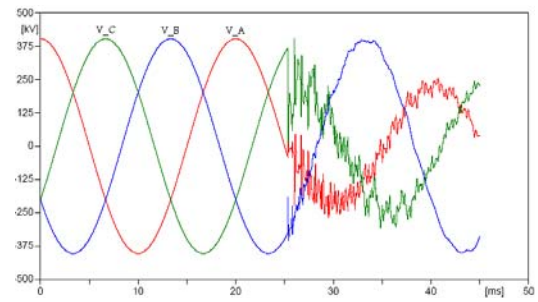

Fig.6 Voltages' waveforms for an ABE fault in 60km distance of receiving busbar at 25 msec.

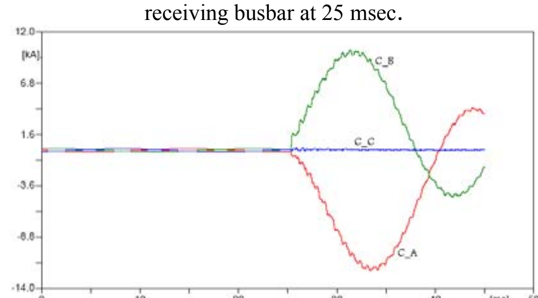

Fig.7 Currents' waveforms for an ABE fault in 60km distance of receiving busbar at 25 msec.

5 Results

The tests of reliability and validity are comprehensive analyzes of the constructed Bayesian and FAM networks that involve both qualitative and quantitative perspectives. Reliability concerns the extent to which a measuring instrument yields consistent results on different trials. Validity concerns the extent to which any measuring instrument measures what it is intended to measure. Reliability is a postulation of validity [13]. In order to criticise the Reliability of the networks, Mean Square Error (MSE) is used and in order to validity of Bayesian network. Table 1 illustrates this criterion.

A typical output and target table of data is shown in Table 1.

Table 1- Typical Output and Targets of Bayesian Network

Item	Phase				Target	Output of Bayesian network
	A	B	C	E		
1	0	0	0	0	1	**1.6315**
2	1	1	1	1	2	2.0765
3	1	0	1	0	4	3.8901
4	0	1	1	0	5	4.9986
5	0	1	1	1	8	8.0896
6	0	0	1	1	11	11.1214

According to the Table 1, a wide range of outputs are similar to the targets, and the Bayesian network has a good results. MSE of results is equal to 0.0563, which is a good operation of this network for fault diagnosis in power systems. Accuracy of this network is about 98.06 percent which is very good for accuracy of diagnosis of faults. The best output of FAM network is a network with vigilance parameter equal to 0.9 and training rate of 0.8. In this network MSE of results is equal to 0.437, and accuracy of this network is about 97.09 percent which is approximately as good as Bayesian network.

6 Conclusion

The analytical effectiveness of a rule-based decision support system relies on its ability to explain its reasoning strategies and results. Indeed, practicing decision analysts find that constructing the qualitative structure of a Bayesian network is more important than precision in numeric parameters. On the other hand, the posterior distribution may then be viewed as the prior distribution of the next iteration in a sequential inference situation. Finally, the interpretation of the Bayesian inference results can be used not only to update the decision-maker's belief about the parameter, but also to conduct a pre-posterior analysis of the constructed Bayesian inference model.

References:

[1] David Heckeman, Abe Mamdani, and Michael P. Wellman, Real-world applications of Bayesian networks, *Communications of the ACM*, Vo1.38, No. 3, 1995, pp. 24-26

[2] T. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufmann, 1988

[3] M. Henrion, J. S. Breese, and E. J. Horvitz, Decision analysis and expert systems, *AI Mag.*, Vol. 12, No. 4, 1991, pp. 64–91

[4] Denison, Holmes, Mallick and Smith, Bayesian methods for nonlinear classification and regression, *published by Wiley*, 2002

[5] Chen-Fu Chien, Shi-Lin Chen, and Yih-Shin Lin, Using Bayesian Network for Fault Location on Distribution Feeder, *IEEE Transactions on Power Delivery*, Vo1.17, No. 13, July 2002

[6] P.Szolovits and S.Pauker, Categorical and probabilistic reasoning in medical diagnosis, *Artif. Intell*, Vo1.1, 1978, pp. 115-144

[7] L. Sovarong and J. S. Costas, A general equipment diagnostic system and its application on photolithographic sequences, *IEEE Trans. Semi-conductor Manufact.*, Vo1.10, 1997, pp. 329-343

[8] R. Duda, J. Gaschning, and P. Hart, Model design in the prospector consultant system for mineral exploration, *in Expert Systems in the Microelectronic Age, D. Michie, Ed. Edinburgh, U.K.: Edinburgh Univ. Press*, 1979, pp. 153-167

[9] W. Buntine, A guide to the literature on learning probabilistic networks from data, *IEEE Trans. Knowl. Data Eng.*, Vo1.8, No.2, 1996, pp. 195-210

[10] Wei Zhao, Xiaomin Bai, Jian Ding, Zhu Fang,Zaihua Li and Ziguan Zhou, A New Uncertain Fault Diagnosis Approach of Power System Based on Markov Chain Monte Carlo Method, *IEEE International Conference on Power System Technology*, 2006, 1-4244-0111

[11] Seyyed Kamal Mousavi, Fault Detection in Transmission Lines by Using Neural Networks, *MSc. Thesis, Amirkabir University of Technology (Polytechnic of Tehran)*, 1998

[12] Aggarwal, R.K., Xuan, Q.Y., Johns, A.T., Furong Li; Bennett, A., A novel approach to fault diagnosis in multicircuit transmissionlines using fuzzy ARTmap neural networks, *Neural Networks, IEEE Transactions*, Vo1.10, No.5, 1999, pp. 1214-1221

[13] E. G. Carmines and R. A. Zeller, Reliability and Validity Assessment, *Newbury Park, CA: Sage*, 1979

[14] Carpenter, G.A. & Grossberg, S., ART2: Self-organization of Stable Category Recognition Codes for Analog Input Patterns, *Applied Optics*, Vo1.26, No. 23, 1987, pp. 4919-4930

[15] Carpenter, G.A. & Grossberg, S., ART3: Hierarchical Search using Chemical Transmitters in Self-organizing Pattern Recognition Architectures, *Neural Networks*, Vo1.3, 1990, pp. 129-152

[16] Carpenter, G.A., Grossberg, S., & Reynolds, J.H., ARTMAP: Supervised Real-time Learning and Classification of Nonstationary Data by a Self-organizing Neural Network, *Neural Networks*, Vol. 4, 1991, pp. 565-588

[17] Carpenter, G.A., Grossberg, S., & Rosen, D.B, Fuzzy ART: Fast Stable learning and Categorization of Analog Patterns by an Adaptive Resonance System, *Neural Networks*, Vo1.4, 1991, pp. 759-771

[18] Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., & Rosen, D.B., Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps, *IEEE Transactions on Neural Networks*, Vo1.3, 1992, pp. 698-713