


# Bayesian Network and Compact Genetic Algorithm Approach for Classifying Partial Discharges in Power Transformers

Pedro H. da S. Palhares<sup>1</sup>  · Cacilda de J. Ribeiro<sup>1</sup> · Leonardo da C. Brito<sup>1</sup> · André P. Marques<sup>2</sup> · Cláudio H. B. Azevedo<sup>2</sup> · José A. L. dos Santos<sup>2</sup>

Received: 7 December 2017 / Revised: 9 May 2018 / Accepted: 14 July 2018 / Published online: 23 July 2018  
© Brazilian Society for Automatics–SBA 2018

## Abstract

This paper presents a statistical learning method capable of classifying the incidence level of partial discharges in power transformers. By using the results from acoustic emission measurements, it is possible to detect the presence of partial discharges inside the equipment, allowing the qualitative health monitoring of the transformer's insulation. Therefore, the use of a Bayesian Network is proposed, combined with a Compact Genetic Algorithm tailored for solving mixed integer programming problems, for discretization of the continuous metrics extracted from acoustic emission measurement. Comparing the results with Multilayer Perceptron Neural Network and Decision Tree and after a suitable amount of runs of the algorithm, it was verified that the Bayesian Networks presented superior results.

**Keywords** Acoustic emission · Bayes procedures · Decision support systems · Feedforward Neural Networks · Genetic Algorithms · Partial discharges

## 1 Introduction

To offer quality of service with minimal cost and without interruption is a goal of electric utilities. For this to occur, the important and expensive equipment, power transformers, work without faults or defects that cause interruptions in power supplying are necessary, because their replacement results in an expensive and long process (Marques et al. 2014).

Several studies in the literature emphasize the importance of developing predictive techniques to identify the defects and incipient faults in power transformers. In Marques et al. (2014), 33 years of monitoring of equipment faults is presented and the importance of the implementation of fault detection methods is exposed. In Wang et al. (2002), statistics of faults in power transformers are presented, as well as diagnostic methods such as the measure of the quality of the oil are described. In Júnior et al. (2015), a review of defects and faults in power transformers and how the existing predic-

tion methods can help to identify them before a catastrophic failure occurs are shown.

The transformer insulation system, particularly the cellulosic material, is the component that is degraded the most and, therefore, defines the useful life of the equipment. In order to avoid the unavailability of the service and to reduce the amount of corrective maintenance, predictive techniques have been developed to detect incipient faults resulting from partial discharges (PD), caused by the decrease in the dielectric strength of the insulation system. Among the currently used techniques, the acoustic emission (AE) measurement stands out as it is noninvasive and enables the localization of PDs (Mohammadi et al. 2006), allowing the identification of the defect and the monitoring of the equipment while it is operating.

For a complete and efficient analysis of the condition of the equipment, computational classification tools can be used in conjunction with the predictive techniques to create expert systems which aids the decision-making process by a human expert. Many references in the literature address the problem of pattern classification of power transformers using measured data. Among them, Tang et al. (2008) and Hao and Cai-xin (2007) work with Dissolved Gas Analysis (DGA), using Bayesian Networks (BN) and Neural Networks (NN), respectively. In Kuo (2009), a Multilayer Perceptron Neu-

---

✉ Pedro H. da S. Palhares  
phpalhares@gmail.com

<sup>1</sup> Federal University of Goiás, Goiânia, GO, Brazil

<sup>2</sup> Enel Distribuição Goiás, Goiânia, GO, Brazil

**Table 1** Description of the data collected by the sensors

Quantity	Description
Time (s)	Instant of time that the activity was detected by the sensor
Amplitude (dB)	The maximum amplitude of the signal during detection of the activity
Energy	Energy accumulated during the detection
Duration ( $\mu$ s)	Duration of the activity
Rise time ( $\mu$ s)	Time between the beginning of the activity and the peak amplitude

ral Network (MLP) is used, trained using acoustic emission measured data to identify the defect by classifying features extracted from real-time AE measurements.

This paper presents an approach using AE measurement to estimate the health of the power transformer, in which a Bayesian Network, combined with a Compact Genetic Algorithm (Harik et al. 1999) adapted for solving mixed integer programming problems (MIPcGA), is trained to classify the state of the equipment. This approach allows the qualitative health monitoring of the equipment while in operation. Therefore, if there are signs of incipient faults, preventive maintenance procedures are scheduled, which, by its characteristics, imply in lower expenses, risks, and wear with customers, especially the possibility of prior communication to consumers on the occasion of the need to temporarily suspend the power supply. The aim of the method proposed in this paper is to provide subsidy to the maintenance staff regarding the appropriate decision-making process on the intervention, or not, in the equipment.

This approach provides some advantages over the former ones: (i) allows to correlate variables, in addition to working with degrees of belief (probabilities), if compared with the NN; (ii) allows the calculation of posterior probabilities, which can provide classifications using a subset of nodes; and (iii) an incorrect result is not ruled out, due to the property exposed in items i and ii.

## 2 Incipient Failures and Partial Discharges

According to Butler-Purry and Bagriyanik (2003), incipient faults are those that slowly develop through a gradual deterioration of the insulating system. Because they are not observed at all instants of time, they are difficult to detect. It is therefore of great importance to early detect these failures, because, over time, when the condition of the equipment is degraded due to electrical, chemical, and thermal effects, incipient faults begin to persist and to intensify in the system (Butler-Purry and Bagriyanik 2003). To that end, important techniques for PDs detection emerged, which, according to Mohammadi et al. (2006) and Azevedo et al. (2009), are small sparks resulted from the degradation of the insulating system

of the transformer and which indicate possible incipient failure.

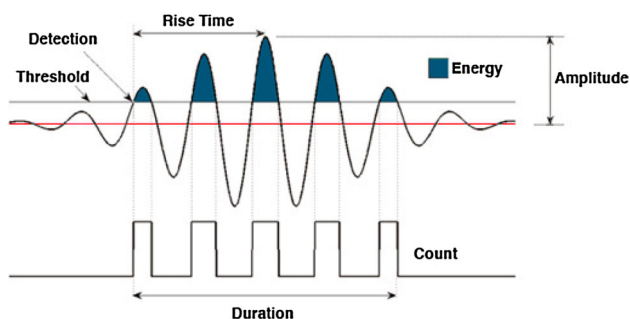
Among the various techniques to detect PDs, in this paper, AE measurement was chosen to use, as it is a noninvasive method that allows the localization of the discharges by monitoring the equipment by using ultrasonic transducers (IEEE 2007), which, positioned strategically and securely on its outer surface, detect any acoustic wave from inside the transformer.

When comparing existing predictive techniques, the dissolved gas analysis (DGA) of the insulating oil, even being an already-established technique in the electricity sector, presents some weaknesses, mainly because of its low sensitivity for detection of partial discharges (IEEE 2007; Azevedo et al. 2009), which does not allow the detection of the incipient faults inside the transformer. In this context, the detection of partial discharges (PDs) by the acoustic emission method (AE) is efficient and also indicates the coordinates of the region where the activities of PDs are occurring (Azevedo et al. 2009). However, as a good practice, it may be considered in conjunction with the DGA method in order to complement the analyst's decision-making process.

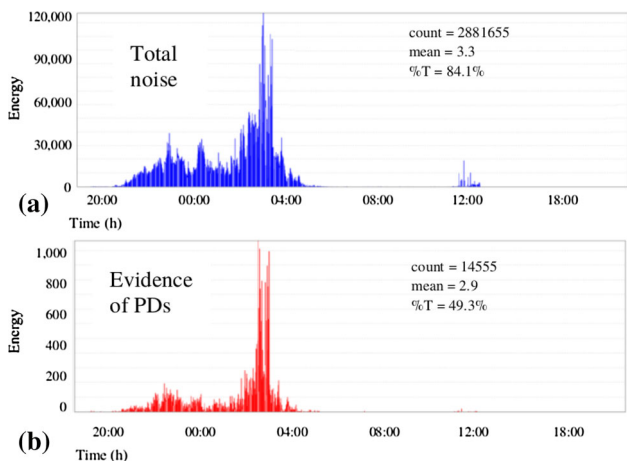
The acquisition of the waveform of the acoustic signal is made optionally by the equipment. Because of the length of each acoustic measurement, which in this work was 24 h (to contemplate the complete cycle of the transformer's load), the storage of the acoustic signal becomes unfeasible, because, grossly, a total of: 14 sensors · 12 bits · 10 Msps · 24 h · 60 min · 60 s  $\approx$  20.5 TBytes would be needed for each equipment, by measurement. Therefore, it was decided to use data obtained from the waveform which are extracted from the parameters from Table 1. On each occurred acoustic activity, only these parameters, not the waveform, are stored in file, which significantly reduces the amount of stored data.

The most important parameters collected by the instrument that records the acoustic signals generated by PDs and collected by the sensors are presented in Table 1. The waveform of a typical acoustic signal waveform is shown in Fig. 1.

In possession of the data provided by the sensors, a procedure is necessary to distinguish between acoustic signals caused by noise and signals caused by PDs. Empirically, it can be noted that the most intense partial discharges occur in the voltage peaks, both positive and negative. It is observed



**Fig. 1** Representation of the activity (occurrence of an acoustic signal waveform) captured by the sensors



**Fig. 2** Variation in energy count as a function of monitoring time, for signals of: **a** total noise, without indication of PDs, and **b** with indication of PDs (Color figure online)

that two activities with a time difference equal to  $1/(2 \cdot 60 \text{ Hz})$  s, equivalent to an angular difference of  $180^\circ$ , can be considered as evidence of partial discharges. It is noteworthy that this calculation must be remade in the case of systems operating in a different frequency of 60 Hz. In addition to this fact, employed sensors have a passband of around 150 kHz, which, for field measurements, is the most suitable range for capturing acoustic signals from partial discharges. In Fig. 2, the results of a measurement are presented with: total noise signals (represented by the blue color) and signals identified with evidence of PDs (represented in red), where count is the number of activities, mean is the average value of energy recorded, and %T is the percentage of minutes where activities were detected.

### 3 Bayesian Networks

Bayesian Networks are Directed Acyclic Graphs (DAG), which nodes have a finite set of states. For each node, there is a table of conditional probabilities that relates each of its values to each possible value from its parents. It is an appro-

priate reasoning model to classify defective equipment, as it allows easy detection of patterns if a suitable set of data is provided (Jensen and Nielsen 2007).

#### 3.1 Bayes Rules

As the foundation of Bayesian Networks, Bayes' theorem is a mathematical tool to calculate conditional probabilities. Its mathematical formula allows the update of unobserved variables probabilities, in light of observed variables. Knowing that  $P(x|e)$  is the probability of  $x$  given the evidence  $e$ ,  $P(e|x)$  the likelihood  $P(x|e)$ ,  $P(x)$ , and  $P(e)$  the prior probabilities of  $x$  and  $e$ , respectively, Bayes' Theorem states that:

$$P(x|e) = \frac{P(e|x)P(x)}{P(e)} \tag{1}$$

#### 3.2 Structure Learning

The process of learning structures of BNs may be done in two ways: (i) by using a prior knowledge of a specialist or (ii) by using an automated tool for structure learning. It is desired to use the latter, as it is an unsupervised way of learning, which allows the detection, at each training, of the best structure for the selected data.

The simplest way to learn the structure of a BN would be to compare all combinations and find which one fits best. For this problem, the exhaustive search becomes impracticable, because the solution space grows super-exponentially as you add nodes (Rudolf Kruse 2013). Therefore, heuristics methods are used to find the best structure for the BN. This paper, therefore, suggests the use of two structure learning methods, K2 (Algorithm 1) and Hill-Climbing (Algorithm 2), selecting the one with the best score.

The K2 (Lerner and Malka 2011) is a greedy search method, which uses its own metric (3) to evaluate the structure. Be  $N_{ijk}$  the number of cases in the training database in which the variable  $u_i$  is instantiated with its  $k$ th value and his parents are instantiated with their  $j$ th value,  $\pi(u_i)$  the parents of  $u_i$ ,  $q_i$  the number of possible instances of  $\pi(u_i)$ , and  $r_{u_i}$  the number of possible discrete values of  $u_i$ :

$$N_{ij} = \sum_{k=1}^{r_{u_i}} N_{ijk} \tag{2}$$

$$g(i, \pi(u_i)) = \prod_{j=1}^{q_i} \frac{(r_{u_i} - 1)!}{(N_{ij} + r_{u_i} - 1)!} \prod_{k=1}^{r_{u_i}} N_{ijk}! \tag{3}$$

where  $g(i, \pi(u_i))$  is the score of the training database given that the parents of  $u_i$  are  $\pi(u_i)$ .

**Algorithm 1** K2

**Input:** Set of network nodes  $U$ , list of examples for training, maximum number of parents.

**Output:** Directed Acyclic Graph.

For each node  $u_i \in U$  do:

1. Find node  $u'_i \in U$  ( $i \neq i'$  and  $i'$  precedes  $i$  in the order of nodes) which maximizes the value of  $g(i, \pi(u_i) \cup \{u'_i\})$ .
2. Add  $u'_i$  to  $\pi(u_i)$ .
3. Continue until you cannot find the variable  $u'_i$  in step 1, or until the number of parents of this node reaches the maximum number.

What makes the K2 attractive is also one of its drawbacks. To run the algorithm, the prior knowledge of the correct order of the nodes is necessary, which is not always possible. Therefore, additionally, the use of the Hill-Climbing algorithm (Jensen and Nielsen 2007) is suggested using the BIC metric (4) and thereby choosing the best structure. Be  $n$  the number of nodes and  $N$  the number of cases in the training base,  $D$  the database of known data and  $G$  the DAG to be scored; the BIC score is given as:

$$\text{BIC}(G, D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_{u_i}} N_{ijk} \cdot \log_2 \left( \frac{N_{ijk}}{N_{ij}} \right) - \frac{\log_2 N}{2} \sum_{i=1}^n q_i (r_{u_i} - 1) \quad (4)$$

The BIC metric has the advantage of being equivalent (Darwiche 2009), i.e., equivalent structures have the same score. Consequently, when generating a new structure, it only requires to calculate the difference between the new and old one, without any need to calculate the score of the entire structure. The valid operations are *add an arc*, *remove an arc*, and *invert the direction of the arc*, always searching for acyclic structures.

**Algorithm 2** Hill-Climbing

**Input:** Initial Structure  $S$  (or consider all nodes disconnected).

**Output:** Directed Acyclic Graph.

**Repeat**

1. Calculate the BIC score for each valid operation in the structure  $S$ .
2. Be  $\Delta^*$  the maximum score obtained in step 1. Be  $A^*$  the operation responsible
3. for  $\Delta^*$ .
4. If  $\Delta^* > 0$ , persist  $A^*$  in  $S$ .

**Until**  $S$  is a local optimum or the number of maximum iterations has been reached.

A third way of constructing a structure is through the use of a naive classifier where all nodes have the same and single parent node.

In this study, two structures were generated, one with each method described (K2 and Hill-Climbing) and the best score was selected as (4). A naive classifier was also employed.

**3.3 Parameters Learning**

The MLE (Maximum Likelihood Estimation) and the EM (Expectation Maximization) are methods for learning parameters. The first operates with complete data, while the latter deals with incomplete data. As the sensors always provide all data related to the activity inside the equipment, the MLE method is recommended, as, according to Darwiche (2009), it is a counting problem, in the form of  $P(X|Y) = N(X, Y)/N(Y)$ , where  $N(X, Y)$  is the number of cases where  $X$  and  $Y$  occur and  $N(Y)$  is the number of cases where  $Y$  occurs.

**3.4 Inference Engine**

Several inference engines in Bayesian Networks are available in the literature (Jensen and Nielsen 2007; Darwiche 2009). This work uses the method of variable elimination, capable of solving exact queries (the probability calculated is exact, not an approximate) of prior probabilities, posterior probabilities, most probable explanation (MPE), and maximum a posteriori probability (MAP), through successive elimination of variables during the process.

**3.5 Discretization of Continuous Values**

The parameterization of the BN requires knowledge of the conditional probability distributions of nodes and their parent nodes, which in the case of continuous variables, would not be a counting problem. Some papers address this problem. Driver and Morrell (1995) use a weighted sum of Gaussians to approximate the prior probabilities (root node) and conditional probabilities, and in Darwiche (2009); Lucas and Hommersom (2015) the continuous variables have to follow a Gaussian distribution.

Therefore, in this work, only discrete values are used. In order to properly and automatically establish the intervals of the parameters, a discretization method is proposed for continuous variables by assigning discrete values, with lower and upper boundaries, to continuous intervals, by using a Compact Genetic Algorithm adapted for solving mixed integer programming problems, the MIPcGA.

**4 Compact Genetic Algorithm for Solving Mixed Integer Programming Problems**

In a traditional Genetic Algorithm (GA), individuals compete against each other and the allele frequency can increase or

**Table 2** Fitness and binary representation of each solution

Solution	Chromosome	Fitness
<i>a</i>	111011	5
<i>b</i>	110100	3

decrease depending on the good or bad decisions made by the GA (Harik et al. 1999). In the cGA, the selection occurs unlike a GA, where genes are evaluated in the context of individuals as a whole. Let us assume two solutions, *a* and *b*, that are competing with each other in an onemax problem, where the objective is to maximize the number of ones in a binary string (Prugel-Bennett et al. 2015).

As presented in Table 2, during the competition, the solution *a* will win by having a greater fitness. The problem in this approach happens in position 4 of the chromosome, where an error occurs by the bit 0 being selected instead of the bit 1. Therefore, Harik et al. (1999) propose a similar schema to a tournament selection: On each iteration, the proportion of winners' genes will grow by  $1/n$ , being *n* the size of the population, benefiting the best genes on each tournament.

By using an appropriate procedure for codification of integer and real variables into bits and a procedure based on Harik et al. (1999) for constraint handling, this paper proposes the use of the cGA tailored for solving mixed integer programming problems, the MIPcGA, for the discretization of continuous variables used in the BN classification process.

The MIPcGA is presented in Algorithm 3. It is noteworthy that between two solutions, the algorithm selects the candidate with a lower constraint violation, or, if they have the same amount, it selects the one with the best evaluation of the objective function. This procedure is similar to Deb (2007). Therefore, individual solutions are compared using the following rules:

1. A solution that violates a smaller quantity of constraints is always preferred over a solution that violates more;
2. Among two feasible solutions, the one having better fitness is preferred.
3. Among two solutions that violate the same quantity of constraints, the one having smaller constraint violation is preferred.

Considering the algorithms proposed, the problem of discretization of continuous parameters of the BN is defined as:

$$\begin{aligned} &\text{Maximize } f(X), \\ &\text{Subject to } l_i \leq x_i \leq u_i, \\ &\quad a_{ij} \in \{0, 1\}, \\ &\quad x_{ij} \in \mathbb{Z}. \end{aligned}$$

**Table 3** Example of valid values of boundaries

Integer variable	Real variable
1	301
0	9874
0	3479
0	1000
1	455

**Table 4** Example of discrete values

Integer variable	Real variable
1	[0, 301]
2	(301, 756]
3	(757, 1756]

where  $f(X)$  is the classification rate of the BN,  $a_{ij}$  represents the bit to enable or to disable the *j*th boundary of the *i*th continuous parameter  $x_{ij}$ ,  $l_i$  and  $u_i$  are the minimum and maximum values of this parameter, and  $X$  is the set containing all  $x_{ij}$ .

Using MIPcGA for discretization of continuous variable implies a solution to a problem of  $2n$  variables, where *n* is the maximum number of intervals. An example of a valid solution is shown in Table 3 and its discrete representation in Table 4, where 0 is the lowest value and 1756 is the highest possible value for the variable. If the integer variable is 0, the real variable is not used to compose the boundaries (it is disabled). If it is one, the real variable composes the boundaries.

To apply the discretization method, the steps below are applied:

1. Learn a BN using an initial solution;
2. Apply the MIPcGA algorithm to enable/disable the boundaries for continuous parameters, selecting genes which maximize the classification rate of the BN. At each generation of the algorithm, new BNs are learned and compared using the selected boundaries.

To evaluate the solutions, in this paper, the steps presented in Algorithm 4 were used to calculate the objective function.

## 5 Power Transformers Classification Using Acoustic Emission Measurements Data

A method is proposed, in this study, for qualitative health monitoring of the power transformers using the procedures described and measurements of acoustic sensors.



**Algorithm 3** MIPcGA

---

**Input:**  $n$ : size of the population,  
 $nr$ : number of real variables,  
 $ni$ : number of integer variables,  
 $nrbits$ : number of bits for real variables,  
 $maxi$ : maximum number of iterations;  
**Output:**  $p$ : final solution.

```

1:  $t_i = \max(maxx - minx + 1)$ ;
2:  $nibits = \lfloor \log_2(t_i - 1) + 1 \rfloor$ ;
3: for  $j = 1$  to  $maxi$  do
4:    $a = \text{Generate}()$ ;
5:    $b = \text{Generate}()$ ;
   (Note that solutions  $a$  and  $b$  compete using the proposed comparison
   criteria,  $bitswin$  represents the bits of the winner, and  $bitloss$ 
   represents the bits of the loser.)
6:    $(bitswin, bitloss) = \text{Compete}(a, b)$ ;
7:    $idx = 0$ ;
8:   for  $j = 1$  to  $ni$  do
9:     for  $k = 1$  to  $nibits$  do
10:       $idx = idx + 1$ ;
11:      if  $bitswin_{idx} \neq bitloss_{idx}$  then
12:         $P_{idx} = P_{idx} + \frac{1}{n}$ ;
13:      else
14:         $P_{idx} = P_{idx} - \frac{1}{n}$ ;
15:      end if
16:    end for
17:  end for
18:  for  $j = 1$  to  $nr$  do
19:    for  $k = 1$  to  $nrbits$  do
20:       $idx = idx + 1$ ;
21:      if  $bitswin_{idx} \neq bitloss_{idx}$  then
22:         $P_{idx} = P_{idx} + \frac{1}{n}$ ;
23:      else
24:         $P_{idx} = P_{idx} - \frac{1}{n}$ ;
25:      end if
26:    end for
27:  end for
28: end for

```

---

**Algorithm 4** Evaluate Solution

- 
- The collected data are split in two groups: 70% for learning and 30% for validation, chosen at random.
  - Using the learning data and considering the boundaries of the solution to be evaluated:
    - Learn the structure of the BN by using the algorithms described in Sect. 3.2 or by using a naive classifier.
    - Learn the parameters by using the MLE method (Sect. 3.3).
  - Check the performance of the BN by measuring the correct classification of the validation data:
    - For each case  $i$  in the validation data:  $score_i = score_i + prob(i_c) + hit(i)$ , where  $prob(i_c)$
    - returns the probability of the correct classification of  $i$  and  $hit(i)$  returns 0 in case of a misclassification and 1 in case of a correct classification.
  - The value of the objective function to be minimized is:  $f(x) = \sum score_i$ .
- 

In order to assist the decision-making process, classifications (from “A” to “D”) were used, with their respective recommended actions, as presented in Table 5 (Marques et al. 2016). The parameters from Table 6 were used in the con-

struction of the BN, which are extracted from data obtained from the measurement (Table 1), as a result of the performed studies described in Marques et al. (2016).

## 6 Results and Discussion

To evaluate the method, 40 measurements performed in power transformers of the company Enel Distribuiçã Goiás were used, with power ranging from 32.0 to 75.0 MVA, with diversified structural designs. The data were collected during AE measurements performed in the field, with the transformer in operation, during 24 h, i.e., a complete load cycle, to ensure all different operating conditions were evaluated.

Fourteen sensors from a DISP device of Physical Acoustics Corporation were used, with a sampling rate of 10 Msps and operating around 150 kHz, positioned at strategic parts of the equipment, requiring full knowledge of the geometry of the design of the active part, tank, and bushings of the power transformer. Each sensor of each measurement was classified, resulting in 560 cases composed by the parameters presented in Table 6. The approach presented in Sect. 2 to distinguish between signals of noise and partial discharges detected by the sensors was used. The optimization algorithm seeks to generate a classification method capable of providing results equal to those given by the specialists who pre-classified the available data set. Therefore, there were 392 samples for training the BNs and 168 samples to validate (70–30%), including all possible classifications, from “A” to “D.” Algorithm 4 was applied 30 times, both to the structure learning (Sect. 3.2) and to the naive classifier (BN-Naive), and for each BN, the performance was verified 100 times, by selecting the learning and validation groups at random. To the continuous values, log 10 was applied, as it reduces the scale of the search space of the MIPcGA optimizer.

By using structure learning, among the 30 BNs created, a correct classification average of 79.94% was obtained, with the average maximum of 85.67% and the minimum of 72.25%. Figure 3 shows all executions of the algorithm. The BN-Naive obtained a correct classification average of 81.82%, with the average maximum of 88% and the minimum of 77.33%. Figure 4 shows all executions of the algorithm using a naive classifier.

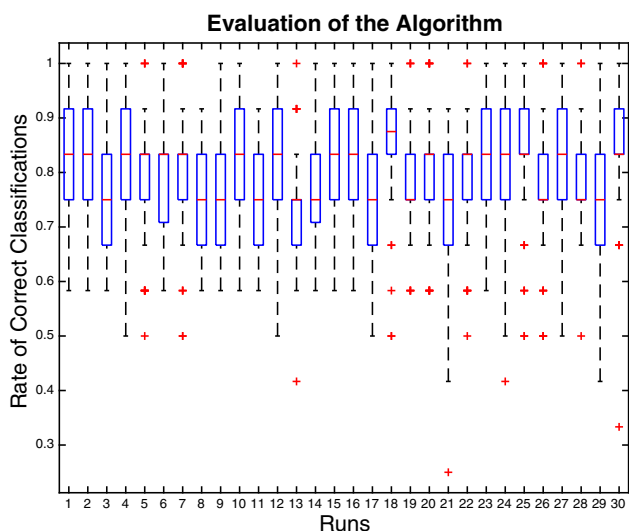
For comparison purposes, Multilayer Neural Network (MLP) trained using the Levenberg–Marquardt algorithm (Mohamad et al. 2010) and Decision Tree (DTree) created with the CART algorithm (Kuhn and Johnson 2013) were used. Different combinations of neurons and layers were tested, for the MLP, and it was found that, for this specific problem, one inner layer with five neurons was the best configuration. The process of classification was run 100 times, and the measurements data were split into two groups, learning and validation, of 70 and 30% of the data, respectively.

**Table 5** Possible classifications

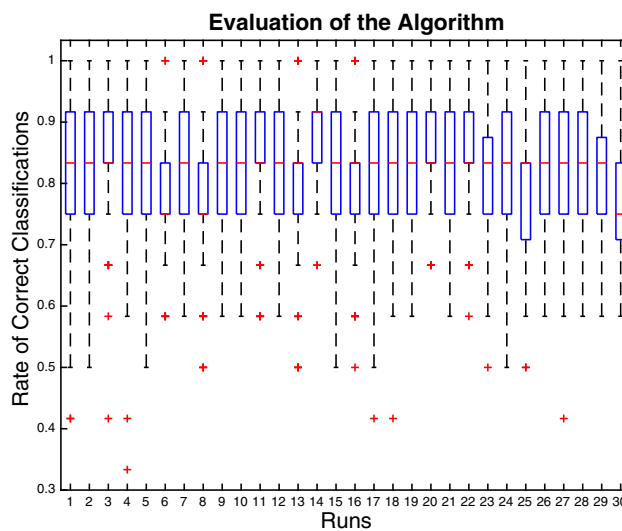
Classification	Action
A	Continue to operate normally
B	Continue to operate normally being attentive to the evolution of PDs in the next measurements
C	Continue the investigation and conduct other tests as soon as possible to confirm the measurements and trend of PDs
D	Plan an emergency removal of the operating equipment for an inner inspection, localization, and correction of the defect

**Table 6** Information used for classification

Node	Description
Count	Number of activities detected by the sensors
Mean	Average value of energy recorded in the measurement
Percentage	The percentage of minutes during the measurement where activities were detected



**Fig. 3** Evaluation of the algorithm using structure learning

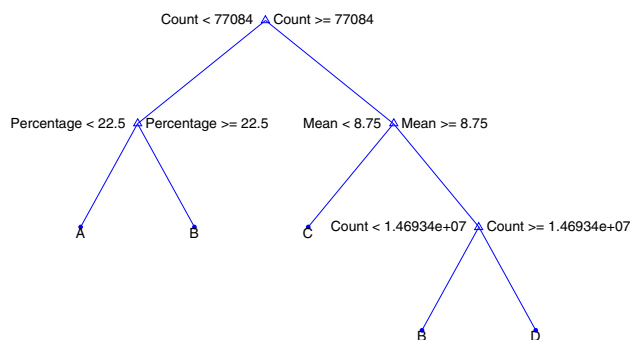


**Fig. 4** Evaluation of the algorithm using a naive classifier

The number of neurons and inner layers of the MLP was varied, confirming that, for this problem, an inner layer with five neurons was the best setting. The correct classification mean of the MLP was 73.67%, with its minimum value of 8.33% and the maximum of 100%. The correct classification mean of the Decision Tree illustrated in Fig. 5 was 77.17%, with its minimum value of 33.33% and the maximum of 100%. The application log 10 in continuous values showed no improvement on the results of the DTree and the MLP.

Table 7 shows a comparison of correct classification rate of the algorithms. To choose the best BN, the intervals which lead to the best average rate of correct classifications were used.

Figure 6a, b shows structures of a naive classifier and the BN using the structure learning algorithm, chosen at random. Tables 8 and 9 show boundaries found by the MIPcGA.



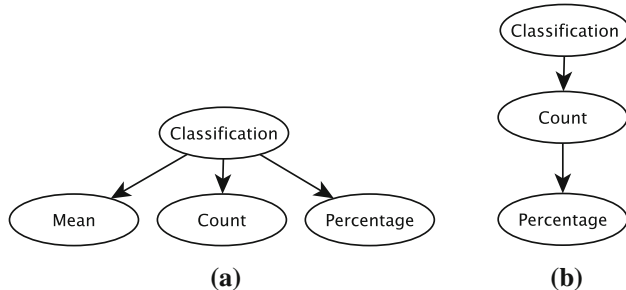
**Fig. 5** Example of a decision tree used for classification

### 7 Conclusion

This paper presented a method capable of learning a suitable Bayesian Network for classification of the health condition of power transformers, considering acoustic emission measure-

**Table 7** Comparison of the classification algorithms

Amount of success (%)	BN (%)	BN-Naive (%)	MLP (%)	DTree (%)
Smallest	50	66.67	8.33	33.33
Average	85.67	88	73.67	77.17
Greatest	100	100	100	100

**Fig. 6** BNs' structures: **a** Naive classifier, **b** structure learned by the algorithm**Table 8** Discrete values of one of the 30 naive classifiers chosen at random

Discrete value	Count	Percentage	Mean
1	[0, 3.9111]	[0, 100%]	[0, 1.55]
2	(3.9111, 6.2815]	–	(1.55, 2.64]
3	(6.2815, 7.4667]	–	(2.64, 9.95]
4	–	–	(9.95, 12.2]

**Table 9** Discrete values of one of the 30 BN using structure learning chosen at random

Discrete value	Count	Percentage	Mean
1	[0, 4.503]	[0, 1.72%]	[0, 1.1]
2	(4.503, 7.4667]	(1.72%, 1.91%]	(1.1, 4]
3	–	(1.91%, 100%]	(4, 12.2]
4	–	–	–

ments data. Such network is able to indicate the classification of the equipment, but also the percentage of the belief in the result, given the evidence. A solver was proposed, for the discretization of continuous values that feed the Network, based on the Compact Genetic Algorithm for solving mixed integer optimization problems with restrictions. The proposed method involves the learning of a BN and the discretization using the MIPcGA. Running a metaheuristic algorithm can be time-consuming; however, the algorithm only needs to be used as the BN needs to be relearned. Therefore, after the BN is learned, quick classifications can be performed without running the solver.

As comparison criteria, Multilayer Perceptron Neural Networks and Decision Trees were created, both are suitable

methods for a classification problem, and it was found that the BN achieved superior performance. Therefore, the proposed method presented better results, both in the worst case, as on average when compared to the MLP and DTree and in the best scenario, showed superior results to the MLP and the same result as the DTree.

The results indicated BN-Naive as the most suitable for the classification process, and then the BN with structure learning. It is noteworthy that the method is suitable for other classification problems, as it can use both discrete and continuous parameters and, also, use structures indicated by an expert, a BN-Naive, or a BN with structure with learning, provided that there is an adequate set of learning data.

**Acknowledgements** The authors would like to thank the Federal University of Goiás and the Federal Institute of Education, Science and Technology of Goiás for their technical support, as well as the Foundation for Research Support of the State of Goiás (Fundação de Amparo à Pesquisa do Estado de Goiás, FAPEG) and the Brazilian Electricity Regulatory Agency (ANEEL) for their financial support of this work.

## References

- Azevedo, C. H. B., Marques, A. P., & Ribeiro, C. J. (2009). *Methodology for the detection of partial discharges in power transformers*. St. Petersburg: EUROCON.
- Butler-Purry, K. L., & Bagriyanik, M. (2003). Identifying transformer incipient events for maintaining distribution system reliability. In *36th Hawaii international conference on system sciences*.
- Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks* (1st ed.). Cambridge: Cambridge University Press.
- Deb, K. (2007). An efficient constraint handling method for genetic algorithm. *Computer Methods in Applied Mechanics and Engineering*, 186, 311–338.
- Driver, E., & Morrell, D. (1995). Implementation of continuous Bayesian networks using sums of weighted Gaussians. In *11th conference on uncertainty in artificial intelligence*.
- Hao, X., & Cai-xin, S. (2007). Artificial immune network classification algorithm for fault diagnosis of power transformer. *IEEE Transactions on Power Delivery*, 22, 930–935.
- Harik, G. R., Lobo, F. G., & Goldberg, D. E. (1999). The compact genetic algorithm. *IEEE Transactions on Evolutionary Computation*, 3, 287–297.
- IEEE. (2007). IEEE guide for the detection and location of acoustic emissions from partial discharges in oil-immersed power transformers and reactors.
- Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs* (2nd ed.). Berlin: Springer.
- Júnior, H. F., Costa, J. G. S., & Olivas, J. L. M. (2015). A review of monitoring methods for predictive maintenance of electric power



- transformers based on dissolved gas analysis. *Renewable and Sustainable Energy Reviews*, 46, 201–209.
- Kruse, R., Borgelt, C., & Braune, C. (2013). *Computational intelligence: A methodological introduction* (2nd ed.). Berlin: Springer.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (1st ed.). Berlin: Springer.
- Kuo, C. (2009). Artificial recognition system for defective types of transformers by acoustic emission. *Expert Systems with Applications*, 36, 10304–10311.
- Lerner, B., & Malka, R. (2011). Investigation of the k2 algorithm in learning Bayesian network classifiers. *Applied Artificial Intelligence*, 25(1), 74–96.
- Lucas, P. J. F., & Hommersom, A. (2015). Modeling the interactions between discrete and continuous causal factors in Bayesian networks. *International Journal of Intelligent Systems*, 30, 209–235.
- Marques, A. P., Azevedo, C. H. B., dos Santos, J. A. L., Souza, F. R. C., Ribeiro, C. J., Moura, N. K., Rocha, A. S., & Brito, L. C. (2016). Metrics for partial discharge detection by the acoustic emission technique to support decision-making. In *IEEE electrical insulation conference*.
- Marques, A. P., Ribeiro, C. J., Azevedo, C. H. B., dos Santos, J. L. A., Sousa, F. R. C., & Brito, L. C. (2014). Power transformer disruptions—A case study. *IEEE Electrical Insulation Magazine*, 30, 17–21.
- Mohamad, N., Zaini, F., Johari, A., Yassin, I., & Zabidi, A. (2010). Comparison between Levenberg–Marquardt and scaled conjugate gradient training algorithms for breast cancer diagnosis using MLP. In *6th international colloquium on signal processing and its applications* (pp. 1–7).
- Mohammadi, E., Niroomand, M., & Rezaeian, M. (2006). *Partial discharge localization and classification using acoustic emission analysis in power transformer*. Busan: INTELEC.
- Prugel-Bennett, A., Rowe, J., & Shapiro, J. (2015). Run-time analysis of population-based evolutionary algorithm in noisy environments. In *Proceedings of the 2015 ACM conference on foundations of genetic algorithms XIII, ACM, New York, NY, USA, FOGA '15* (pp. 69–75).
- Tang, W. H., Lu, Z., & Wu, Q. H. (2008). A Bayesian network approach to power system asset management for transformer dissolved gas analysis. In *Third international conference on electric utility deregulation and restructuring and power technologies*.
- Wang, W., Vandermaar, A. J., & Srivasta, K. D. (2002). Review of condition assessment of power transformers in service. *IEEE Electrical Insulation Magazine*, 18, 12–25.