# An industrial Fault Diagnosis System based on Bayesian Networks

Abdelkabir Bacha LISER Laboratory ENSEM Casablanca Hassan II University (UH2C) Ahmed Haroun Sabry LISER Laboratory ENSEM Casablanca Hassan II University (UH2C) Jamal Benhra LISER Laboratory ENSEM Casablanca Hassan II University (UH2C)

# ABSTRACT

This paper presents a DC motor fault diagnosis system based on Bayesian networks. This was done by the design of a new electromechanical test bed allowing the collection of functioning data from a real world industrial Direct current (DC) Motor. The data collection will help in the construction of Bayesian networks models. These data are collected from sensors measuring different types of variables that are directly related to the industrial system. Without doing any mathematical modeling that describes the physical properties of the studied DC motor, the proposed tool provides with the help of Bayesian networks parameters and structure learning algorithms, the base to construct a fault diagnosis tool that can be extended to a fault prognosis tool.

# Keywords

Machine Learning, Artificial Intelligence, Bayesian networks, fault diagnosis, data acquisition, DC motor

# 1. INTRODUCTION

In industrial firms, the availability and reliability of the main production elements or even in the other auxiliary systems are necessities. That's why those firms invest considerable budgets in order to maintain their production systems up in good conditions. In this perspective, fault detection and diagnosis are taking more care and attention [5]. So, the use of artificial intelligence (AI) in this field is more than welcomed. AI is a vast domain, this work will not explore all its fields but the study will be limited in one of its subdomains known by Machine Learning (ML) [18]. Into ML there is what is called probabilistic graphical models [15] or to use another denomination one would say: Bayesian networks (BN) or belief networks [9].

The fault diagnosis that was invoked in the first paragraph can not be done easily unless a considerable knowledge of the system is well gathered. If that is the case, the prediction of the system's behaviour is affordable. In order to do this prediction, a modeling is absolutely needed. Generally, two alternatives exist: One of them consists of trying to provide a mathematical model with some complex equations that are reflecting the system's behavior with of course the help of many simplification hypothesis. Within those hypothesis some phenomenons that that are difficult to model are neglected. This old method gives relatively good results, however it is generally not easy to find those equation [5]. This approach is called model-based approach [20]. The second methodology called data-oriented does not deal with the internal behavior knowledge of any element in the system. It can regenerate a simulated behavior regarding the external excitation of the entire system without any mathematical model. This type of methods treats the studied system as a black box that its internal construction is unknown. It relies on the record and the use of massive data collected from the system [20]. In this stage AI come with the help. More specifically ML in AI that is the more concerned [18] [12]. Those collected data will be used after for doing the learning phase to some algorithms that will predict the researched behavior of the system [15] [4]. Though, the acquisition and the availability of data to be used for ML formalisms such as Artificial Neural Networks and Bayesian Networks are widely posed problems [13] [11].

This paper's approach consist of using one of the latest methods in the modeling domain: A data-oriented method. witch consist of using Bayesian networks or probabilistic graphical Model (another name) [9] [15] [16]. The goal is to test this formalism in a real world system that is widely used in industrial machinery and also in everyday life use: The DC motor. It is a system whose mathematical model is available [13] [1]. So it will be a good real world testing platform to validate the utility of our approach.

In this perspective a test bed equipped with a data acquisition and processing system was especially designed. This is a study of a Direct current (DC) motor with its control system in order to collect the maximum information by connecting different types of sensors that measure various physical quantities: the supply voltage, current, power consumption, many temperature sensors dispatched into different hot-spots, the speed, the duty ratio of the control, vibration, torque developed, the load resistant torque, etc...

# 2. MACHINE LEARNING (ML) & BAYESIAN NETWORKS (BN)

ML is an artificial intelligence discipline that deals with the development, the analysis and the implementation of methods that allow a machine to improve itself through a process of learning, and so perform tasks that are difficult or impossible to fill by conventional algorithmic methods. Bayesian networks are part of this discipline [18]. Bayesian networks were developed to address the difficulties of decision [11] where there is always the presence of uncertainty, for example the detection of presence or absence of a cancer in an organ based on test results and biopsies [9] [14] [15].

#### 2.1 Learning in Bayesian Networks

2.1.1 Parameters learning. This is the kind of learning that is more or less similar to that used for the neural networks [19]. We seek here to estimate probability distributions (or parameters of the relevant laws) from available data. These distributions are often called CPD or CPT (conditional probability distributions) or (conditionnal probability Tables). The most used methods for Bayesian networks learning, according to the kind of available data, and to their completeness (no values are missing in records) are : statistical learning , Bayesian learning for complete data, while learning with incomplete data uses specific algorithms such as the Expectation Maximisation (EM) algorithm [16].

2.1.2 Structure learning. The aim here is to find the structure that better represent our problem. So learning the structure of a Bayesian network from data means finding a graph representing a perfect map (P-map) of an independence model associated with a probability distribution that we have a sample from. So one must be sure of the existence of such P-map (loyalty) and to know all the variables (causal sufficiency) [16]. Two major families of approaches exist: firstly, the search approaches of conditional independence. Secondly, by score based methods [16].

#### 3. THE PROPOSED ACQUISITION SYSTEM

In order to demonstrate the utility and the validity of the use of Bayesian networks in the fault diagnosis in industrial systems, this work joins an antecedent work that have been conducted in [2] concerning the application of this formalism to diagnose failures of an induction Motor. In this previous work, we relied on the knowledge of an expert in the maintenance of such machines to model the relationships between the different nodes of the Bayesian network used. And this was done by using the qualitative method proposed in [19]. This methodology with its benefits, still cannot give reliable results because the values used are human qualitative judgments. To improve this, the present work provides a tool that is based on reliable data extracted directly and in real time with all the sensors associated with the studied system. Therefore, 2 DC motors embedded into a test bench have been chosen in this study.

#### **3.1** Description of the data acquisition system

In Figure 1 a synoptic model for the acquisition system is given. This model contains all components that are responsible for generating the required data. Those components are :

- -The studied DC motor
- —The H-bridge
- -The H-bridge control
- -The current sensor
- -The voltage sensor
- -The temperature sensor
- -The acquisition card

In the following subsections, a detailed description of every component is given.

*3.1.1 The studied DC Motors.* These are high performance DC machines whose technical characteristics are shown in Table 1. They are designed to work in harsh conditions.

Two machines of this mentioned type are used in this work. They are axially coupled. One of them is used in motor mode and the other in generator mode.

 Table 1. Technical specifications of the

 studied DC Meters

studied DC Motors					
Nominal supply voltage	12V				
Operational supply voltage rang	4.8 14.4 V				
Rotation speed	17 200 rpm				
Current at maximum efficiency	7A				
Maximum efficiency	77,00%				
Weight	220 g				

*3.1.2 The H-bridge.* It is an electronic card managed by an integrated circuit (L298N) specialized in driving DC and stepper motors. It is based on the principle of H-bridge that allows the control of the direction and the rotation speed of the motor. One can drive with this card two dc motors or one stepper motor.

3.1.3 The H-bridge control. It is a device for controlling the speed and direction of the motor studied. This is achieved using the pulse width modulation (PWM), which is a periodic rectangular digital signal whose time duration at logic level "1" is proportional to a ratio called the duty cycle and is proportional to the desired speed.

3.1.4 The current sensor. It is a current measuring module based on the Hall Effect sensor (ACS712). this element is symbolized by an ammeter with a dotted arrow to indicate the fact that the current value measured by the sensor is not directly transmitted to a display, but this value is sent to the system electronically via analog to digital conversion (ADC) to be recorded after treatment.

3.1.5 The voltage sensor. The voltage measurement is provided by the acquisition card itself, more precisely by the analog to digital converter module (ADC) integrated in the microcontroller of the card. And since the measured voltages in the these experiments are slightly higher than 5V (the nominal voltage of the motor used is about 12V), a voltage divider based on resistors to transform the reading range to be between 0 and 5V (maximum allowable voltage of the module the microcontroller) has been added. This bridge will decrease the voltage read by multiplying it by a coefficient less than 1. Then, the microcontroller software will multiply the values read by the analogue-digital converter by the inverse of the voltage divider ratio for recording the exact values.

3.1.6 The temperature sensor. It is a widely used digital temperature sensor that uses the one wire protocol to transmit its values to a microcontroller. It is used often in general purpose electronic projects. In this work, tree of this type of sensors are dispatched near an important and functional hot-spots such as the nearby of collectors and brushes of DC motor.

*3.1.7 The acquisition card.* The acquisition card is the heart of the acquisition system. It is based on a microcontroller that handles many tasks such as : the reception of data from sensors, the analog to digital conversion and the transfer of data via the I2C bus and serial console to the unit that will handle these data. The microcontroller is clocked at a frequency of 20 MHz to run those tasks at the same time, in addition to a faster data transfer.

### 4. DATA PREPROCESSING AND THE BN MODEL DEVELOPMENT

#### 4.1 Received Data

All variables that have been successfully measured from the system are:



# Acquisition System Model

- Electrically Controlled Variable Load Load
- H Bridge : Electronic circuit that applies a voltage across the motor in either direction
- : Electronic unit that generates PWM (Pulse Width Modulation) signal to control the Motor speed and direction Control
- µC Based Acquisition Card : Acquisition card based on a microcontroller that transforms the analog signals from sensors to digital ones.

Computer running an OS : Any computer running an Operating System.

Fig. 1. Acquisition System model

- -The current of the studied DC motor
- ---Voltage of the studied DC motor
- -The internal temperature: near brushes and collector.
- —The external temperature.
- -The temperature of the control bridge (H-bridge)
- -The duty cycle: the speed control
- -The working condition of the machine: start = 1: the machine running, start = 0: otherwise.

At the output of the acquisition card, all records of instantaneous states of cited variables are available via the two transmission modes: serial port (UART) and I2C. These data have been collected during varied functioning states (normal operating, with/without load, with overload, with blocked shaft, etc.). Then, they have been registered in extension log files (.txt and .csv). In this stage these data are not suitable for use and need to be cleaned and well filtered.

#### 4.2 Data preprocessing

The Bayesian networks modeling process begins with a very important task: filtering and cleaning data. This is called feature engineering [12]. Since the major number of variables in data acquisition concerns continuous variables, it is necessary to discretize them in order to be able to use them in the next phase [6]. Previous works such as [17] and [7] proposed a methodology in this field. For this case, after obtaining a dataset from the system, it was cleaned by deleting manually some aberrant values. Then, a 10 Bins unsupervised discretization filtering algorithm was applied for every continuous attribute. This algorithm is available in the opensource software called Weka 3.6 [10] [8]. As for Boolean attributes the same algorithm was applied but this time with only 2 Bins and a Make Binary option in order to prepare the ground to the next learning algorithms.

# 4.3 Learning the structure and parameters of the model

4.3.1 Learning using the K2 Algorithm. In order to learn the structure and parameters of the BN, The K2 algorithm proposed by [3] was used. This algorithm works only with nominal and no missing values only. It can be initiated in manner that outputs a Naïve Bayes structure for simplification purposes. It has also the possibility to add a Markov Blanket correction to the learned classifier. This ensures that all nodes in the network are part of the Markov blanket of the classifier node. Another important parameter of this algorithm concerns the number of parents that each node can have.

4.3.2 Learning the structure of a BN using K2 initiated as Naïve Bayes. Firstly, the K2 algorithm was initiated as Naïve Bayes learner so that the initial structure will be a Naïve Bayes Network. A Naïve Bayes network is a network with an arrow from the classifier node to each other node. If this option is not used, an empty network will be used instead as an initial network structure. The number of parents per node was set to 1. Figure 2 shows the structure of the learned network as the described conditions above says, the Figure shows also the learned conditional probability tables (CPT) per each node and per each value of the node.

4.3.3 Learning the structure of a BN using K2 algorithm initiated as Naïve Bayes with max number of parents greater than 5 for each node. After learning the first BN, a change of some parameters of the algorithm is suitable in order to find an optimal configuration that gives a better score functions. So, the parameter called MaxNbOfParent that designates the maximum number of parents that itch node in the graph can have was varied. Figure 3, shows the learned BN after choosing the right number of parents per each node which is exactly the number 5. After various iterations we found that the graph converges to this structure even if we increase the said parameter for values greater than 5.

### 5. RESULTS & DISCUSSIONS

#### 5.1 The chosen criterion for learning the structure of Bayesian Networks

The K2 algorithm was configured to choose as an optimisation function the LogScore bayes, this function must be maximized [10]. The score type determines the measure used to judge the quality of a network structure. It can be one of : Bayes, Minimum Description Length (MDL), Akaike Information Criterion (AIC), Entropy, Etc.

Table 2.	10 folds	cross v	validation	results for	the K2 algorithm
----------	----------	---------	------------	-------------	------------------

	Initiated as Naive Bayes		
	N = 1	N >=5	
Correctly Classified Instances	1307 (99.8472 %)	1309 (100 %)	
Incorrectly Classified Instances	2 (0.1528 %)	0 (0 %)	
Kappa statistic	0.9951	1	
Mean absolute error	0.0016	0	
Root mean squared error	0.0383%	0.0008 %	
Relative absolute error	0.5149 %	0.0147 %	
Root relative squared error	9.6934 %	0.2083 %	

Table 4. Confusion matrix for both graphs

N	N = 1		>= 5	
a	b	а	b	<- classified as
252	2	254	0	a = '(-inf-0.5]'
0	1055	0	1055	b = '(0.5-inf)'

#### 5.2 Performance of the learned Bayesian Networks

5.2.1 The evaluation method. The presented models in Figures 2 and 3 were learned using a cleaned and discretized dataset of 1309 instances. This dataset has 12 attributes; five of them have 10 values while the rest are binary. In order to judge the performance of those models, many methods exist in the literature. We are using one of the most used performance analysis methods in machine leaning that gives reliable results: the K-folds cross-validation applied to classification of faults using this dataset. Exactly, we are using 10 folds in test mode.

5.2.2 Comparison of two generated graphs. In order to compare between the first generated BN structure that was initiated as Naïve Bayes with maximum number of parents per each node is equal to "1" (N = 1), and the other one that was initiated also as Naïve Bayes but this time with maximum number of parents that each node can take greater than 5 (N >5), Tables 2 gives the 10 folds cross validation results for both graphs. The maximum number of parents were denoted by N. In this table a comparison is given between the two graphs regarding the use of these models in classification tasks. The first graph succeeded to correctly classify 1307 instances by 1309 of total instances (two instances were misclassified) achieving 99.84% of the classification correctness and 0.0016 of Mean absolute error. The second BN graph is more performing than its predecessor achieving 100% of classification correctness and 0 of Mean absolute error.

5.2.3 Detailed accuracy for both graphs. As for Table 3, it shows a detailed Accuracy by class. The class term means the classification variable "Fault" that has two values. The first value is "(-inf-0.5]" : a denotation of the used software [10] [8] meaning that the probability of the classified instance is between [0 and 0.5]. The second value "(0.5-inf)" means that the probability of the classified instance is into the interval ]0.5, 1]. And finally, for both graphs, a confusion matrix is given in Table 4.

5.2.4 ROC Curves. In Figure 4 Receiver operating characteristics or ROC curves are given for boths graphs to be compared with an ROC curve for the Zero R classification algorithm witch is a tree based classification algorithm that was applied to the same dataset of 1309 instances. In those plots one can see that both curves that correspond to the two learned Bayesian Networks are nearly identical and give areas under curves nearly equal to "1" witch is a good result, and that they are dominating the Zero R ROC curve witch gives a poor performance according to this figure.



Fig. 2. Learned Naïve Bayes BN structure with learned CPT per each node using the K2 algorithm

		Table 5.	uctaneu /	ceuracy by ci	ass		
	TP Rate		FP Rate		Precision		Class
	N = 1	N >=5	N = 1	N >=5	N = 1	N >=5	
	0.992	1	0	0	1	1	(-inf-0.5]'
	1	1	0.008	0	0.998	1	'(0.5-inf)'
Weighted Avg.	0.998	1	0.006	0	0.998	1	
	Recall		F-Measure		ROC Area		Class
	N = 1	N >=5	N = 1	N >=5	N = 1	N >=5	
	0.992	1	0.996	1	1	1	(-inf-0.5]'
	1	1	0.999	1	1	1	'(0.5-inf)'
Weighted Avg.	0.998	1	0.998	1	1	1	

Table 3. detailed Accuracy by class

# 6. CONCLUSION & PERSPECTIVES

In this work, the utility of Bayesian networks in the diagnosis and detection of failures in industrial systems was presented. Then, a methodology was proposed for learning Bayesian networks from a data streaming received in real time from an acquisition system that was especially designed by the authors for this purpose. A detailed description was given to all composing elements of this acquisition system. Then, a presentation was given of the exploitation of these data for learning Bayesian networks generated for the detection, diagnosis or even prognosis of failures that may arise on the studied system. In fact, these data was cleaned, because some of the received values from the system were continuous. This was done by applying a dicretization algorithm on those variables in order to make them useful for the learning algorithms. After these steps, two Bayesian Networks were learned, one of them was initiated as a Naïve Bayes structure with no ability for its nodes to have more then 1 parent, while the other one was initiated as a Naïve Bayes

structure with the freedom to every node to have more then one parent.

#### 7. REFERENCES

- E Afjei, A Karami, et al. Sensorless speed/position control of brushed dc motor. In *Electrical Machines and Power Electronics*, 2007. ACEMP'07. International Aegean Conference on, pages 730–732. IEEE, 2007.
- [2] Abdelkabir BACHA, A.Haroun SABRY, and Jamal BENHRA. Contribution l'aide la décision dans le domaine industriel en utilisant les réseaux bayésiens. In CIMSI 14 Conférence Internationnale sur le Monitoring des Systmes Industriels.
- [3] Gregory F Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial intelligence*, 42(2):393–405, 1990.



Fig. 3. Learned BN structure with CPT for each node using the K2 algorithm initiated as Naïve Bayes and with max number of parents per each node greater than 5



Fig. 4. ROC curves of the two learned graphs compared to a Zero-R classification algorithm ROC curve applied to the same Dataset

- [4] Gregory F Cooper and Edward Herskovits. A bayesian method for the induction of probabilistic networks from data. *Machine learning*, 9(4):309–347, 1992.
- [5] Marcos FSV DAngelo, Reinaldo M Palhares, Luciana B Cosme, Lucas A Aguiar, Felipe S Fonseca, and Walmir M Caminhas. Fault detection in dynamic systems by a fuzzy/bayesian network formulation. *Applied Soft Computing*, 21:647–653, 2014.
- [6] M Julia Flores, José A Gámez, Ana M Martínez, and José M Puerta. Handling numeric attributes when comparing

bayesian network classifiers: does the discretization method matter? *Applied Intelligence*, 34(3):372–385, 2011.

- [7] Nir Friedman, Moises Goldszmidt, et al. Discretizing continuous attributes while learning bayesian networks. In *ICML*, pages 157–165, 1996.
- [8] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. ACM SIGKDD explorations newsletter, 11(1):10–18, 2009.
- [9] David Heckerman, Dan Geiger, and David M Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995.
- [10] Geoffrey Holmes, Andrew Donkin, and Ian H Witten. Weka: A machine learning workbench. In *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*, pages 357–361. IEEE, 1994.
- [11] Khalid Iqbal, Xu-Cheng Yin, Hong-Wei Hao, Qazi Mudassar Ilyas, and Hazrat Ali. An overview of bayesian network applications in uncertain domains. *International Journal of Computer Theory and Engineering*, 7(6):416, 2015.
- [12] Patrick Jahnke. *Machine Learning Approaches for Failure Type Detection and Predictive Maintenance*. PhD thesis, tudarmstadt, 2015.
- [13] Andrew KS Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510, 2006.

- [14] Kiran R Karkera. Building Probabilistic Graphical Models with Python. Packt Publishing Ltd, 2014.
- [15] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques.* MIT press, 2009.
- [16] Philippe Leray. Réseaux bayésiens: apprentissage et modélisation de systèmes complexes. PhD thesis, Université de Rouen, 2006.
- [17] Stefano Monti and Gregory F Cooper. A multivariate discretization method for learning bayesian networks from mixed data. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 404–413. Morgan Kaufmann Publishers Inc., 1998.
- [18] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [19] Patrick Naïm, Pierre-Henri Wuillemin, Philippe Leray, Olivier Pourret, and Anna Becker. Réseaux bayésiens. *Paris: Eyrolles*, 1999.
- [20] Xiao-Sheng Si, Wenbin Wang, Chang-Hua Hu, and Dong-Hua Zhou. Remaining useful life estimation–a review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1):1–14, 2011.