On line diagnosis of gas turbines using probabilistic and qualitative reasoning

Zenón Flores–Loredo, Pablo H. Ibargüengoytia
Instituto de Investigaciones Eléctricas,
Calle Reforma 113, Cuernavaca,
Morelos, 62490, México,
zfl,pibar@iie.org.mx.

Eduardo F. Morales
ITESM, Campus Cuernavaca,
Av. Paseo de la Reforma 182-A, Temixco,
Morelos, 62589, México,
eduardo.morales@itesm.mx

***Abstract*— Artificial Intelligence (AI) techniques are becoming an active area of research for real applications. Power industry is one of the best examples of this. Different problems have been solved with these techniques, for example monitoring, alarm management, diagnosis, and network planning. This paper presents an On-line diagnosis system for gas turbines in power plants. Since this application deals with unexpected behavior, probabilistic reasoning and specifically Bayesian networks, offer a natural mechanism for diagnosis. However, the use of Bayesian networks in real applications presents two challenges. First, the acquisition of representative models of the process with and without faults. Second, dealing with continuous variables makes very expensive the computation for inferences. This project utilizes automatic learning algorithms, together with expert advice to determine the models of the most common faults in gas turbines. Also, a quantification of the behavior is used to minimize the cost of the probability propagation in Bayesian networks. These produces an original probabilistic and qualitative diagnosis of gas turbines. Experiments were carried out utilizing real data in a simulator. The results are presented and discussed.**

I. INTRODUCTION

Power generation is considered a complex and expensive process. In these days, the power generation industry has faced important problems that require the modernization of current installations, principally in both the instrumentation and control systems. The current trend consists in increasing the performance, availability and reliability of the actual installations. The performance refers to the amount of mega watts that can be generated per unit of fuel. The availability refers to the hours that the central stops generating electricity, and the reliability refers to the probability of counting with all the equipment of the plant running. Additionally, modern power plants are following two clear tendencies. First, they are working very close to their security limits in order to obtain the maximum efficiency and hence, the best performance. Second, they are highly automated and instrumented, leaving the operator with very few decisions. However, the classic control systems are programmed to stop the turbine under the presence of abnormal behavior, even if the fault may be weak. Control systems are unable to make decisions in some cases. However, at the supervisory level it is possible to $reason$ about the abnormal behavior, the probable causes and the consequences. This is the role of a diagnostic system.

Diagnosis is the technique utilized in several fields, devoted to find faults that explain abnormal behavior in a system. Several approaches have been proposed and they can be classified in three kinds [1]:

- data–driven: based on large amount of data provided by modern control and instrumentation systems, from which meaningful statistics can be computed.
- analytical: based on mathematical models, often constructed from physical first principles.
- knowledge–based: based on causal analysis or expert knowledge, where conclusions and inferences are made given information of the process. They can be found in several kinds of models and inference methods [2].

The selection of the best approach for a given problem depends on the quality and type of available models, and on the quality and quantity of data available.

This research team, at the Electrical Research Institute or IIE, has been working in the design of on–line intelligent diagnosis systems for gas turbines of power plants [3], [4], [5]. This kind of projects include two special problems. First, the management of uncertainty given the thermodynamic conditions of the gas turbine and the difficulty of constructing accurate analytical models of the process. Second, the continuous acquisition of the turbine parameters for their analysis. Solving these problems will allow the early detection of small deviations of the expected behavior.

This paper presents a qualitative and probabilistic diagnosis system for gas turbines. Three special challenges are solved in this project. First, the definition of the most common faults in the different stages of the operation of the turbine, namely starting up, normal generation of power, and special maneuvers. Second, the acquisition of probabilistic models of the gas turbine, given real data from the plant and special advice from the experts. Third, since most of the signals that conform the models are continuous value variables, then a practical treatment is required. Then, the main contribution of this work can be established as follows: It presents a complete methodology for the construction of real applications of Bayesian networks. This methodology consists in the discretization of variables to their qualitative tendencies, the model induction, and their integration in an architecture for on-line diagnosis.

This paper is organized as follows. First, section II introduces the probabilistic model of the dynamic behavior of the turbine, as well as the techniques utilized to obtain those models. Section III describes the application domain, and explains the characteristics of the faults considered in this work. Next, section IV presents the experiments carried out and the experimental results obtained in the laboratory. Finally, section V concludes the paper and addresses the future work in this area.

II. Probabilistic Diagnosis Model

This paper presents the utilization of Bayesian networks in the diagnosis of gas turbines, i.e., the early detection of a deviation of the normal behavior given the evidence obtained in real time.

Bayesian networks are directed acyclic graphs representing probabilistic dependence and independence relationships between the variables in a domain. Nodes represent the variables and the arcs represent the probabilistic dependence of the variable at the end of the arc, from the variable at the beginning of the arc. Evidence is entered when the value of some nodes exists. Thus, different algorithms allow the propagation of probabilities in order to obtain the posterior probability of a variable, given the evidence [6].

A typical approach would include a structure like the one shown in Fig. 1. In the upper level, external factors are included representing events that can produce a fault. For example in medicine, contact with sick people can cause the contagion. In the lower level, foundings or symptoms are included. For example, presence of fiber or headache. Given evidence of one or all of these factors allows to propagate the probabilities and *infer* certain fault or sickness.

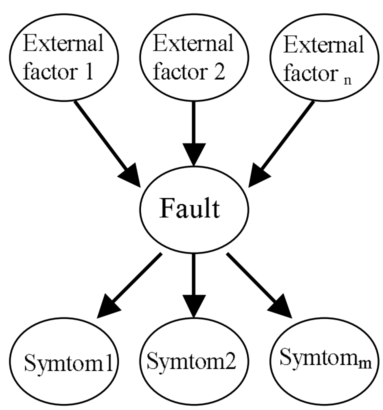

Fig. 1. Typical approach in the construction of diagnosis systems using Bayesian networks

However, a strong disadvantage of this approach is the difficulty to obtain accurate models with accurate parameters. Thus, this project explored the utilization of automatic learning algorithms that provide the probabilistic models, given data obtained from a real plant or using a gas turbine simulator.

Figure 2 shows the basic software architecture developed in this project. Data from the gas turbine simulator are utilized to create data files that are used by the learning process. In this research, the K2 induction algorithm for Bayesian network is used. K2 [7] is a score-based greedy search algorithm for learning Bayesian networks from data. It was published by Cooper and Herskovits in 1992. K2 uses a Bayesian score, $P(Bs, D)$, to rank different structures and it uses a greedy search algorithm to maximize $P(Bs, D)$. Inputs to K2 include: a set of nodes, an ordering on the nodes, an upper bound u on the number of parents a node may have, and a database D containing m cases of training data. For each node, K2 returns its most probable parents given the training data D. This is carried out off–line.

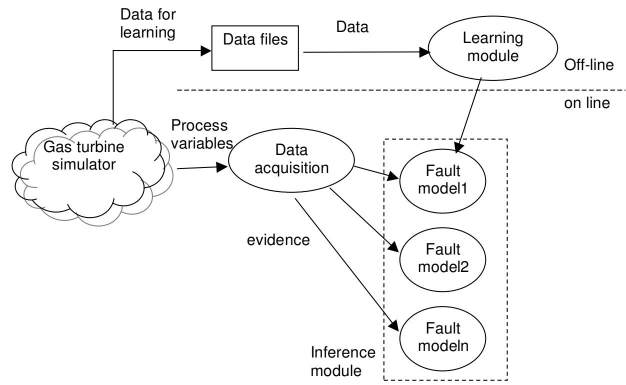

Fig. 2. Proposed software architecture for multiple faults diagnosis of gas turbines.

Different models are obtained for different faults. This schema assumes that all the faults are independent and assumes that multiple simultaneous faults can be detected independently, i.e., they do not cancel one another.

The inference module is formed by independent threads that receive evidence from the data acquisition module. This is carried out on–line, i.e., data are refreshed every 250 ms. Thus, the calculation of the probability of the different faults are calculated also on–line and reported to the operator.

Given the dynamic operation of the turbines, two stages are considered independently. In the first phase, the starting up of the plant. In the second phase, the generation of power at full load. In the first, the heating and the increment of speed of the turbine are the main operations. The second stage is the normal generation of power at different charges. In both stages, different faults can be detected. The next section explains the gas turbines and their possible faults in more detail.

III. Application domain: gas turbines

The gas turbine is one of the most important equipment in a combined cycle power plant. Figure 3 shows a simplified schema of the gas turbines at *Dos Bocas* and *Gomez Palacio* power plants in Mexico.

A gas turbine consists fundamentally of four main parts: the compressor, the combustion chamber, the turbine itself and the generator. The compressor feeds air to the combustion chamber, where the gas is also fed. Here, the combustion

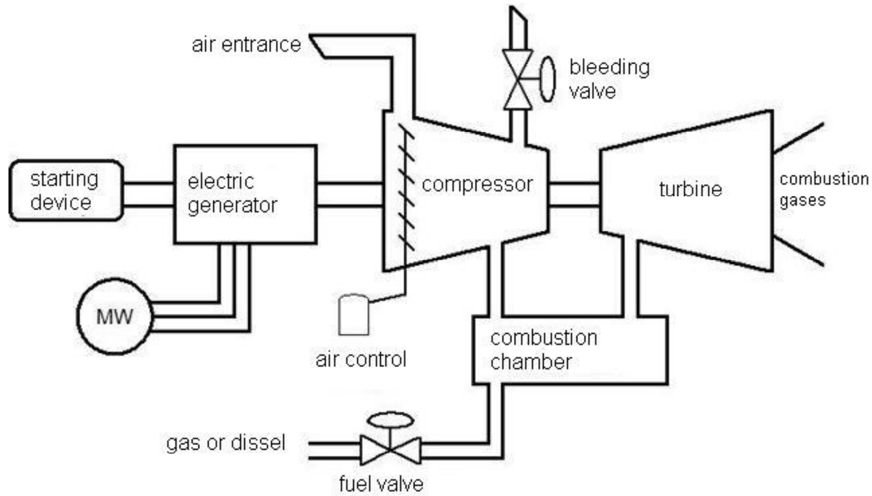

Fig. 3. Simplified schema of a gas turbine.

produces high pressure gases at high temperature. The expansion of these gases in the turbine produces the turbine rotation with a torque that is transmitted to the generator in order to produce electric power output. The air is regulated by means of the *inlet guide vanes* (IGV) of the compressor, and a control valve does the same for the gas fuel in the combustion chamber. The control valve is commanded by the control system or by the operator in manual operation mode, and its aperture can be read by a position sensor. The temperature at the blade path, which is the most critical variable, is taken along the circumference of the turbine. Other important variables, measured directly through sensors are the mega watts generated and the turbine speed in revolutions per minute (rpm).

The initial experiments carried out in this project include the following faults:

- Low fuel supply pressure: The fuel provider presents low pressure in the supply. The control calculates the aperture of the gas valve based on a certain supply pressure, so incorrect aperture is commanded and the fault occurs.
- Fault in the compressor bleed valves: theses valves stabilize the pressure of the turbine in the start up phase. The pressure may increase or decrease to dangerous levels if the valve is stuck at incorrect values.
- Permanent stuck of the fuel valve: the fuel valve gets stuck and has no response to the control commands.

The fuel valve malfunction has three main sources. First, excess of friction in the actuator of the valve causes an increment in the difference between the real position and the commanded position. Second, the position of the valve remains stuck in some value. Third, the valve may unstuck unexpectedly and cause a large amount of flow of fuel. This may cause an increment of temperature that the control algorithm will try to compensate, with the final result in a dangerous unbalance of the process and an oscillation of temperature. More detail with other faults can be consulted in [8].

IV. Experiments

Data for the experiments were obtained utilizing a gas turbine simulator at the IIE laboratory. Table I shows the set of variables utilized and their identifier.

TABLE I
VARIABLES PARTICIPATING IN THE EXPERIMENTS.

IDentifier	Description
Vtg	turbine speed
RVtg	turbine speed reference
MW	generated power
RMW	demand of power generation
XVc	valve of gas position
RXVc	reference of valve of gas position
EXVc	error in the valve of gas position
Pcc	combustion chamber pressure
Piqg	gas pressure in the burners
scag	IGV position signal
Rscag	demand of IGV position signal
XVs	compressor bleed valve position
Ipp	main switch position
Marr	starter engine state

During operation, data are collected every 250 ms. with the value of the variables before the simulated fault, and some samples of data after the simulated fault. This data is given to the automatic learning module and different models for different faults are obtained. One experiment corresponds to the start up phase of the turbine, in which the speed goes from 0 to 5100 rpm without load Other experiment corresponds to the power generation phase. The simulation includes the command of full load, going from 2 MW to the maximum load of 24 MW, and then returning to 2 MW.

Practically all the signals in table I are continuous variables. Using a representation based on $traditional$ Bayesian networks, a discretization process is required. However, if a large number of intervals is chosen, then a great amount of memory and computer time is required to obtain the posterior probabilities. An additional problem when using a large number of intervals is that many configurations are not represented in the database, and so the corresponding probabilities cannot be estimated, thus reducing the accuracy of the model. If a short number of intervals is chosen, then lack of expressivity is obtained. In this work, a qualification of the variables is proposed, similar to the human reasoning performed in the control rooms, i.e., operators observe if a signal increases, decreases or remains unchanged. For example, the operator knows that at the start up phase, if the temperature increases, then the speed must also increase. Thus, the nodes of the network can have one of three states, according to the relation shown in table II.

TABLE II

DESCRIPTION OF THE QUANTIFICATION PROCESS.

Condition	Variable state
if $v_t - v_{t-1} > \delta$	s_{up}
if $v_t - v_{t-1} < \delta$	s_{down}
if $v_t - v_{t-1} = \pm\delta$	s_{rem}

where v_t represents the value of a variable at current time, v_{t-1} represents the previous value, and δ represents a small threshold dependant on a percentage according to the noise in the signal.

Using this qualification process, data obtained in the simulator are transformed and applied to the automatic learning process.

Figure 4 shows the resultant Bayesian network obtained with the K2 algorithm of the $Elvira$ package [9] for the fuel valve fault.

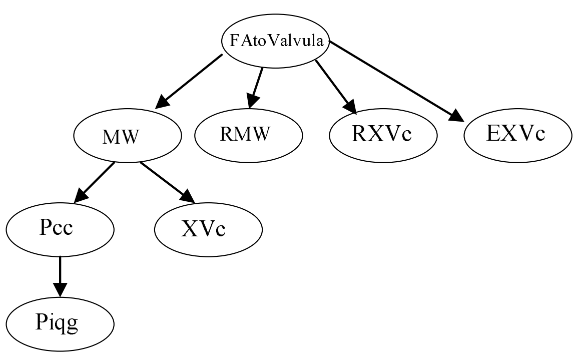

Fig. 4. Network obtained automatically by applying the K2 algorithm for fuel valve fault.

Similar models were obtained for different faults as described in Fig. 2. Table III shows the initial results. The rows indicate the case study applied to the prototype. The columns describe the results obtained for the different fault models given the same evidence.

For example, in the first row, experiments with the simulator in the start up phase without simulated fault, produces only a probability of 15 % of fault in the compressor bleed valve. This percentage is considered very low, indicating no fault. In the last 4 rows, a specific fault was simulated and reported.

TABLE III

PRELIMINAR RESULTS OF THE EXPERIMENTS.

Case	Fault 1	Fault 2	Fault 3
start up, no fault	No detected	15 %	No detected
generating, no fault	No detected	No detected	No detected
simulating fault 1	100 %	No detected	No detected
simulating fault 2	No detected	98 %	No detected
simulating fault 3	No detected	No detected	95 %

Additional tests are being developed in the laboratory, including variations for the other faults.

Figure 5 shows the photo of the prototype user interface. At the right side, the state of the process and the presence of faults are indicated. The gas turbine may be in the start up (**arranque**), stop (**paro**), synchronization (**sincronizacion**) or generating (**generación**). Also, the probability of fault is shown in colors: green for no fault (low or null probability), yellow for medium probability (warning) and red for high probability of fault. At the bottom, windows displays the probability of faults with numerical value, and the exact time when the fault occurred.

V. CONCLUSIONS AND FUTURE WORK

This paper has presented the design and testing of an intelligent diagnosis system for gas turbines. The system uses the models of the abnormal behavior of the turbine, given certain fault. The models are represented by Bayesian networks considering a binary node representing the presence of the fault. The Bayesian networks were induced through the K2 Bayesian network learning algorithm, using data obtained from a simulator. In the networks, all the nodes are continuous variables. This paper proposes the quantification of the signals in three values: signal going up, down or stable.

Four faults were included in the experiments, from two stages of the operation of the turbine. The preliminary results are promising but further experiments are needed, utilizing the gas turbine simulator in the laboratory. Future work include experiments in a real power plant, considering real data.

Qualitative probabilistic diagnosis has demonstrated to be an appropriate mechanism for the on–line diagnosis of real processes.

REFERENCES

[1] L. Chiang, E. Russell, and R. Braatz, *Fault Detection and Diagnosis in Industrial Systems.* London, Great Britain: Springer, 2001.

[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "Review of process fault diagnosis." *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 293–346, 2003.

[3] P. H. Ibargüengoytia, L. Sucar, and S. Vadera, "Real time intelligent sensor validation," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 770–775, 2001.

[4] L. de Jesús González-Noriega and P. H. Ibargüengoytia, "An architecture for on–line diagnosis of gas turbines," in *Advances in Artificial Intelligence - IBERAMIA 2002, LNAI 2527*, M. T. F.J. Garijo, J.C. Riquelme, Ed. Sevilla, Spain: Springer, 2002, pp. 795–804.

[5] B. Morales and P. H. Ibargüengoytia, "On–line diagnosis using influence diagrams," in *MICAI 2004: Advances in Artificial Intelligence, LNAI 2972*, L. S. H. S. R. Monroy, G. Arroyo, Ed. México D.F., México: Springer, 2004, pp. 546–554.

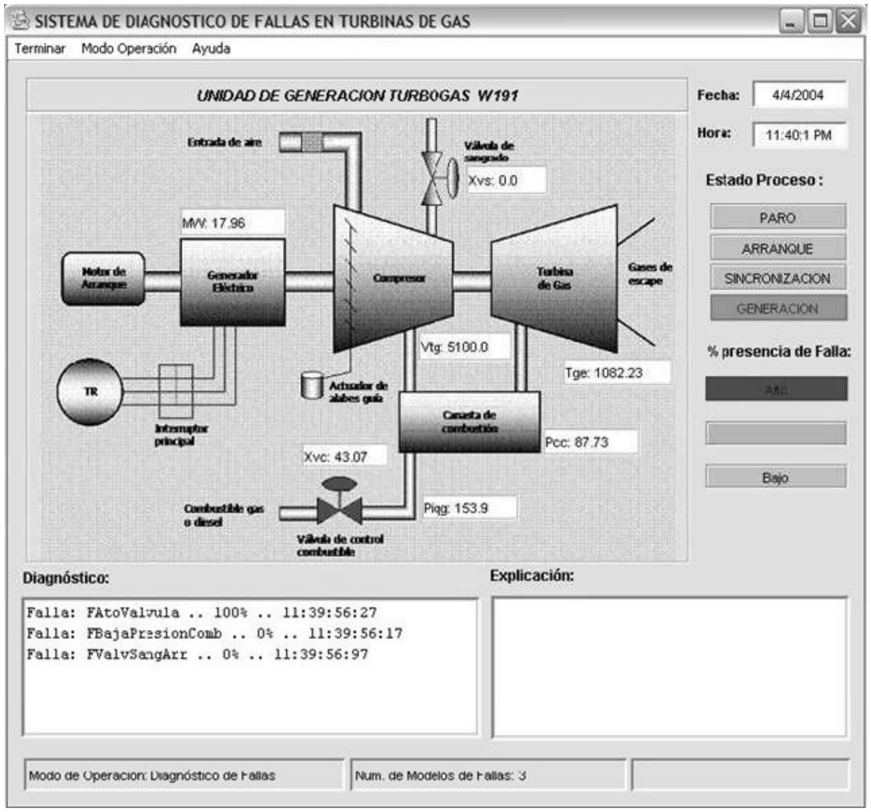

Fig. 5. Interface of the prototype (in Spanish).

[6] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Palo Alto, Calif., U.S.A.: Morgan Kaufmann, 1988.

[7] G. Cooper and E. Herskovitz, "A bayesian method for the induction of probabilistic networks from data." *Machine Learning*, vol. 9, no. 4, pp. 309–348, 1992.

[8] M. A. Delgadillo, "Descripción de fallas en turbinas de gas," Instituto de Investigaciones Eléctricas, IIE," Technical Report, 2004.

[9] T. Elvira Consortium, "Elvira: An environment for creating and using probabilistic graphical models," in *Proceedings of the First European Workshop on Probabilistic graphical models (PGM'02)"*, Cuenca, Spain, 2002, pp. 1–11.