SEDRET—an intelligent system for the diagnosis and prediction of events in power plants

G. Arroyo-Figueroa, Y. Alvarez, L.E. Sucar

Abstract

Artificial Intelligence applications in large-scale industry, such as fossil power plants, require the ability to manage uncertainty and time. In this paper, we present an intelligent system to assist an operator of a power plant. This system, called SEDRET, is based on a novel knowledge representation of uncertainty and time called Temporal Nodes Bayesian Networks (TNBN), a type of Probabilistic Temporal Network. A set of temporal nodes and a set of edge define a TNBN, each temporal node is defined by a value of a variable and a time interval associate to the change of variable value. A TNBN generates a formal and systematic structure for modeling the temporal evolution of a process under uncertainty. The inference mechanism is based on probabilistic reasoning. A TNBN can be used to recognize events and state variables with respect to current plant conditions and predict the future propagation of disturbances. SEDRET was validated with the diagnosis and prediction of events in a steam generator with a power plant training simulator. The results performed in this work indicate that SEDRET can potentially improve plant availability through early diagnosis and prediction of disturbances that could lead to plant shutdown.

Keywords: Diagnostic expert systems; Knowledge-based systems; Temporal probabilistic networks; Fossil power plant application

1. Introduction

There is a strong tendency to design supervision and control systems with artificial intelligence techniques (Boyen & Wehenkel, 1999; Kang & Golay, 1999; Zhang & Zhao, 1999). Current economic, social and environmental factors put stringent requirement on steam power plants to be operated at high level of efficiency and safety at minimum cost. The result has been an increase in the complexity of power control system operations (Arroyo-Figueroa, Sucar, Solis & Villavicencio, 1998). In a steam power plant, an operator has to monitor several hundred measurements and alarms. Under fault situations, the operator of the plant must be able interpret each measurement that is received by the control system, and determine which is the condition of the equipment in order to make a proper control action. The complexity of the decisions that the operator is required to make is continually increasing along with the severity of the consequences of an error in judgement. In addition, the ability to respond quickly can often be the decisive factor in the prevention of the developing malfunction. While their abilities may match the demands of day-to-day operations, the flood of alarms and upset indications generated by a disturbance (process malfunction or fault) can overwhelm even the most rigorously trained operators.

A fossil power plant can be described by great variety of processes with multiples state variables, events and disturbances. In this domain, the state variables change over time in response to both internal and external disturbances as well as the transition of time itself. In the process is a signal exceeding its specified limit of normal functioning called an event, and a sequence of events that have the same underlying cause are considered as a disturbance. During disturbances, the operator must determine the best recovery action according to the type and sequence of the signals received. Current control systems do not provide the means for intelligent interpretation of sensor data, diagnostic problems, coping with large process disturbances or predicting the consequences of control action (Arroyo-Figueroa & Villavicencio, 1994; Falinower & Mari, 1994; Moradian, Thompson, Tomlinson & Jenkins, 1992; Shirley, Forbes & Nelson, 1990; Tsou, 1995; Wong, Ho & Teo, 1994).

In this paper, we present an intelligent system to assist an operator of fossil power plants. The intelligent system called Intelligent System for the Diagnosis and Prediction of Events (SEDRET) is a Shell based on probabilistic networks.
to deal with uncertainty and time. The research goal is to develop a real-time knowledge based system to act as an operational and diagnostic aid to operators of power plants.

The paper is organized as follows. First, we show how the probabilistic network can deal with uncertainty and time, and present a novel methodology called Bayesian network (BN) with temporal nodes (TNBN). Afterwards, we present the design and development of an intelligent system to assist an operator of fossil power plant based on a TNBN. Finally, we illustrate its application in a fossil power plant with a detailed example: the diagnosis and prediction of events in the drum system of a steam generator.

2. Temporal nodes Bayesian network

In the large-scale industry, such as fossil power plants, the process information is generally imprecise and incomplete, and it changes over time. These problems suggest that any successful representation should handle uncertainty and time in a principled and unambiguous way. An ideal representation should be sound and complete, facilitate efficient inference, as well as be amenable to explanation methods. We propose a novel representation based on a probabilistic network for dealing with uncertainty and time, called Temporal Nodes Bayesian Network (TNBN) (Arroyo-Figueroa & Sucar, 1999).

Bayesian networks are a robust and sound formalism to represent and handle uncertainty in intelligent systems in a way that is consistent with the axioms of probability theory (Pearl, 1988). A BN is a graphical structure (Directed Acyclic Graph) composed of nodes and arcs, used for representing expert knowledge. In this graphical structure, each node corresponds to an entity of the real world (variable: hypothesis or evidence) and each link gives direct information about the dependency relationships between the variables involved. These dependency relationships are parameterized by conditional probabilities required to specify the underlying distribution. In particular, the qualitative knowledge is represented by the topology of the network and the quantitative knowledge is represented by the joint probability distribution of the variables. A BN has a number of important properties, including a declarative semantics as an extension of probability theory, efficient and complete inference algorithms, and several machine learning and explanation methodologies. There are several exact and approximate inference algorithms for BNs, and special case algorithms and corresponding conditions that allow tractable inference. The inference mechanism is based on probabilistic reasoning. This consists of instantiating the input variables, and propagating their effect through the network to update the probability of the hypothesis variables. However, BNs were not designed to model temporal relationships between the process variables (Aliferis & Cooper, 1996; Santos & Young, 1996). The main problem is to represent each variable with its interactions with

Fig. 1. Temporal nodes Bayesian network for the drum level system.
### Table 1
Variable definition and a priori and conditional probabilities (time in seconds)

<table>
<thead>
<tr>
<th>Probabilities a priori for the node “Feedwater Pump Failure (FWPF)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occur</td>
</tr>
<tr>
<td>Does not occur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probabilities a priori for the node “Feedwater Valve Failure (FWVF)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occur</td>
</tr>
<tr>
<td>Does not occur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probabilities a priori for the node “Load Increment (LI)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occur</td>
</tr>
<tr>
<td>Does not occur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional probabilities for the node “Feedwater valve increase (FWV)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>FWVF</td>
</tr>
<tr>
<td>True [28–41]</td>
</tr>
<tr>
<td>True [41–66]</td>
</tr>
<tr>
<td>False [28–66]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional probabilities for the node “Feedwater pump current augmentation (FWP)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>FWP</td>
</tr>
<tr>
<td>LI, [10–29]</td>
</tr>
<tr>
<td>True, [29–107]</td>
</tr>
<tr>
<td>False, [10–107]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional probabilities for the node “Feedwater Flow Increment (FWF)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>FWV</td>
</tr>
<tr>
<td>True [114–248]</td>
</tr>
<tr>
<td>False [25–248]</td>
</tr>
<tr>
<td>FWF</td>
</tr>
<tr>
<td>True [144–248]</td>
</tr>
<tr>
<td>False [10–135]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional probabilities for the node “Drum level high condition (DHL)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHL</td>
</tr>
<tr>
<td>True [10–27]</td>
</tr>
<tr>
<td>True [27–135]</td>
</tr>
<tr>
<td>False [10–135]</td>
</tr>
</tbody>
</table>
other variables over multiple points of time (Horvitz & Seiver, 1997). We propose an extension of the probabilistic semantics to deal with temporal relations.

2.1. Definition of a TNBN

In a simple probabilistic network, each node represents a variable and each edge is a dependency relationship between two nodes. In the case of a temporal nodes BN, each node represents a change of the variable, denominated as an “event” and each event is associated to a time interval of occurrence (Arroyo-Figueroa, 1999). Hence, each edge represents a causal and temporal relationship between two temporal nodes. In our approach a time interval is an additional component of the probabilistic network. Formally each temporal node is defined as:

Definition 1. A temporal node (TN) is a set of ordered pairs \((s, \tau)\), where \(s\) is the state or value variable and \(\tau\) is the time interval associated with each state variable attribute. \(\Sigma\) represents the set of state variable and \(T\) represents the set of time intervals. There is a default state that corresponds to no change (generally the “normal” state) with a temporal range of interest as time interval.

Temporal nodes are connected by edges. Each edge represents a causal and temporal relationship. The dependency relation between two nodes is affected by the time interval of occurrence. The conditional probability distribution for each node is defined as the probability of each ordered pair \((s, \tau)\) given the ordered pairs of its parents \((s, \tau)\). Formally a TNBN is defined as:

Definition 2. A TNBN is defined as \(TNBN = (V, E)\), where \(V\) is the set of temporal nodes and \(E\) is the set of edges. Each temporal node is defined by an ordered pair \((s, \tau)\) and the conditional probability matrix that specifies the probability of each ordered pair given its parents.

A consequence of this model is the increase of the states of the node needed to represent the state changes of the process. This means an additional number of the assignments in the joint probability distribution. The TNBN can be see as a natural expansion of a probabilistic network, hence the properties of TNBN are parallel to the properties of a probabilistic network. The inclusion of time intervals into the network is unconstrained. However, it is necessary to make a trade off between the temporal expressiveness and the complexity. An advantage of this kind of representation is the ability to diagnose and to predict events, and their time interval of occurrence, based on the occurrence of certain events using standard propagation techniques developed for probabilistic networks. This consists of instantiating the input variables (events occurrence) and propagating their effect through the network to update the probability of the hypothesis variables. When an event is detected, the reasoning mechanism is used to update the marginal probability of all the nodes. The next section presents an example, how the TNBN formalism can be use for modeling the process evolution of a fossil power plant.
2.2. An example of a TNBN

In this section we give an example of using a TNBN model to represent and apply temporal knowledge. Fig. 1 presents a small TNBN that captures part of the causal and temporal knowledge about an industrial process. Table 1 gives the definition of the TNBN of the example and the conditional probabilities associated with each temporal node. The model is based on the following knowledge obtained from an expert operator of a power plant about the drum level system operation:

The drum is a tank with a steam valve at the top, a feedwater valve at the bottom, and a feedwater pump, which provides water to the drum. When a power load increase occurs the current in the feedwater increases, this take several seconds after the power load increase. Another cause of current augmentation of the feedwater pumps is when a pump failure occurs, but this take less seconds. Both disturbances produce an increase in the feedwater flow inside the drum tank. This will lead to an increase in drum level to a dangerous level, this take some seconds. The operator must open the steam valve in order to increase the steam flow. This will lead to a reduction of the water drum level in the drum tank so that the level will decrease to safe levels. Another cause of increase of the feedwater flow is the failure of the feedwater valve, this takes several seconds after the feedwater valve increase of its opening. This also will lead to an increase in drum level to a dangerous level but this increase takes more time.

2.3. Evaluation of a TNBN

The inference mechanism of a TNBN model is based on the detection of the events and the propagation of the evidences. The inference mechanism updates the marginal posterior probabilities of each node (variable) of the network given the occurrence of an event or events. We define tc as the time when an event is detected and \( \alpha \) as the real time occurrence function, this function is defined as the absolute value of the difference between the time of occurrence of a pair of connected events. As the net does not have any temporal reference, the time of occurrence of the first event fixes temporally the network. The value of \( \alpha \) is used to determinate the time interval of the “effect” node considering the “cause” node as initial event. Afterwards, the evidence is propagated through the network to update the probabilities of the other nodes. These probabilities
show the potential occurrence of the past and future events. The stop condition is when a terminal or leaf node is reached.

We can identify three main steps in the inference mechanism procedure:

1. detection of the event or events occurrence and definition of the time interval of occurrence of the event(s);
2. propagation of the evidence occurrence through the net and update of the probabilities of the variables;
3. determination of the potential past and future events.

2.3.1. Step 1: event detection and time interval definition

When an initial event is detected, its time of occurrence, $t_{\text{initial}}$, is utilized as temporal reference for the network. There are two possible cases, depending on the position of the initial node in the network: (a) the initial event corresponds to a root node, (b) the initial event corresponds to an intermediate or leaf node.

2.3.2. Step 2: propagation of the evidences

Once the value of a node is obtained (time interval and associated state), the next step is to propagate the effect of this value through the network to update the probabilities of other temporal nodes.

2.3.3. Step 3: determination of the past and future events

With the posterior probabilities, we can estimate the potentially past and future events based on the probability distribution of each temporal node. If there is not enough information, for instance there is only one observed event that corresponds to an intermediate node, the mechanism handles different scenarios. The node is instantiated to all the intervals corresponding to the observed state, and the posterior probabilities of the other nodes are obtained for each scenario. These scenarios could be used by an operator or a higher level system as a set of possible alternatives, which will be reduced when another event occurs, see Arroyo-Figueroa (1999) for details.

3. SEDRET

SEDRET is a computer tool for building, validating and running Temporal Nodes Bayesian Networks (Alvarez-Salas, 1999). SEDRET can be used to recognize the significance of the events and state variables in relation to current plant conditions and predict the future propagation of disturbances. Structurally, SEDRET has six modules: knowledge base builder, knowledge base, probabilistic-temporal inference engine, plant operation database, data acquisition, and

Fig. 4. Zoom of the sub-system (drum) where the event was detected.
operator interface. The computer tool allows the semi-automatic building of a TNBN. It also facilitates the validation of the TNBN. The process values of the variables of the domain can be read from keyboard (off line operation) or requested by the system (on line operation). The system allows on-line application in a plant. SEDRET has been implemented in Visual C 5.0 and Visual Basic under Windows 98 on a PC Pentium III. The architecture of SEDRET is shown in Fig. 2.

3.1. The knowledge base

A TNBN represents the process knowledge in SEDRET. The knowledge base in SEDRET is organized by objects. Each node and arc of the net is an object defined by its attributes. A node is an object that represents an event or variable and an arc is an object that represents the causal and temporal relationships between two nodes. The attributes of each node are name, description, position (root, intermediate and leaf) and probability parameters. The attributes of each arc are the connections between the nodes ("cause" and "effect").

3.2. Knowledge base builder

The goal of the knowledge base builder is to help in the construction of the probabilistic temporal net. Hence, the knowledge database builder has an object editor, for the events and arcs. The object editor defines the network structure: number of nodes, description of each node, relations between the nodes, and probability parameters. The knowledge database builder has a friendly man–machine interface, see Fig. 1.

This module also shows the information and properties of each node and arc in the TNBN. It facilitates the analysis and validation of the information of each object in the net. Once that the probabilistic network is built, it is possible to modify the structure, by adding or eliminating nodes. In a similar manner, it shows the information with regard to the operative state of a node when SEDRET is in the on-line operating mode.

3.3. Probabilistic inference

The real-time inference engine in SEDRET reasons about the current state of the process. The reasoning mechanism implemented within SEDRET is an approach for probability propagation in multiply connected networks called propagation in trees of cliques (Pearl, 1988). The inference engine operates when an event is detected and it defines the time interval of occurrence and the state of the temporal node,
propagating the effect through the net. Once the knowledge base has been built, the knowledge engineer can then start the reasoning process. The probabilistic inference engine updates the marginal probabilities of the nodes in the network. These probabilities are used to show the most probable cause of an event and the possible future events (prediction).

3.4. Plant operation database

SEDRET has a database of the process variables and equipment of the plant. The database consists of six workspaces: water–steam generator, feedwater system, condenser system, steam–turbine system, superheater system and reheater system. Each of these plant workspaces is a collection of variables related to its plant area. The attributes of each variable are its name, description, nominal value, and related variables.

3.5. Operator interface

This module has been designed for displaying the results of the diagnostic and prediction tasks. The user interface receives information from the process data and the inference engine and displays the results on the screen. This module helps the user follow the state of each variable and the events detected. The user interface has been designed in order to facilitate the dialogue between operators and the system. Fig. 3 shows the main windows of SEDRET when an event is detected, the abnormal variable is highlighted in red. Afterwards, the user clicks on the highlighted component to proceed with the analysis of the causes and effects.

3.6. Data acquisition module

This module allows data to be obtained from one of the two sources: the data acquisition interface (on line operation) or the keyboard interface (off line operation). SEDRET has two operating modes: off-line and on-line. In the off-line mode, the process data are introduced through a keyboard. In the on-line mode, the process information is obtained from a data acquisition system of a power plant simulator or a real plant. The power plant simulator emulates the dynamic behavior of the critical state variables, such as flow, pressure, and temperatures of the process.

4. Experimental results

SEDRET has been applied for fault diagnosis and prediction of events in a steam generator of a fossil power plant. We consider the water–steam generator system with four
potential disturbances: a power load increase (LI); a feedwater pump failure (FWPF); a feedwater valve failure (FWVF) and a spray water valve failure (SWVF). The water–steam generator is a system that provides superheated steam to the steam–turbine system and takes the water of the condenser system.

In the process, a signal exceeding its specified limit (high or low) is called an “event”, and a sequence of events that have the same underlying cause are considered as a “disturbance”. To determine which of the disturbances is present is a complicated task, because there are similar sequence of events for each of the disturbances. We need additional information in order to determine which is the real cause. In particular, the temporal information about the occurrence of each event is important for an accurate diagnosis. For example, a feedwater flow increase (FWF) can be produced by two different events: a feedwater pump current augmentation (FWP) and a feedwater valve opening increase (FWV). We can use the time difference between the occurrence of each event, FWV–FWF and FWP–FWF, for selecting the “cause” of the increase of the FW flow.

The network structure was defined based on the knowledge of an expert operator, see Fig. 1. The definition of the time intervals for each temporal node was obtained based on the knowledge about the process dynamics combined with the data from a simulator. Once the structure and time intervals were defined, the required parameters were estimated from the data. A training simulator of a 350 MW fossil power plant generated the process data. We selected 80% of this database (800 registers) for parameter learning and 20% (200 registers) for evaluation.

The evaluation is performed as follows. The selected disturbance is first simulated on a fossil power plant training simulator (MICROTHERM 300). During these simulations, SEDRET is actively performing its analysis and the relevant real-time performance is recorded. At the same time, all the data from the simulator that SEDRET needs for its analysis are sent to SADRET. The analysis begins when an event is detected.

### 4.1. First experiment

The drum is a subsystem of a fossil power plant that provides steam to the superheater and water to the wall of a steam generator. One of the main problems in the drum is to maintain the level in safe operation. We assume that a feedwater flow increase (FWF) is detected at 4:38:00. This time is called the occurrence time of the event (tcFWF). The related variables are feedwater valve (FWV), feedwater pump (FWO) and drum level condition (DHL). The status
of these variables is “normal” only with a positive increase for the case of FWP. The SEDRET screen in Fig. 4 shows a zoom of the system and Fig. 5 shows the relevant information about the detected event.

The analysis of the propagation of the event is shown in the SEDRET screens in Figs. 6 and 7. Fig. 6 shows the probability value (certainty value) of the initial cause of the event and the probabilities of the past and future events and its time interval of occurrence. If the FW flow occurs in the first time interval the initial cause is the FW pump failure (the velocity of the pump tends to maximum) with a probability value (certainty) of 0.75. The recommended actions are (1) run the stand bye FW pump and (2) repair the FW pump. The recommended actions are in the Plant Operation Database and they are update by the analyzer system. The probable past events are a FW pump increase, 25–114 s before the FW flow event, and FW valve opening increase, 114–248 s before the FW flow event, with a probability of 0.92. The probable future event is a high level condition, 29–107 s after the FW flow event detection, with a probability of 0.95.

This experiment shows how the time difference between the occurrence of the events is useful for reducing the uncertainty to increment the quality of the diagnosis.

4.2. Second experiment

The second experiment shows the effect of a FW flow increase in the superheater system of the steam generator. The superheater is a system that provides superheated steam to the main turbine and takes saturated steam from the drum system. One of the main problems is to maintain the steam temperature at an optimal value. Temperature control is considered to be the most demanding control in the steam generator. The steam flow, the drum conditions (FW flow and Steam flow) and the spray water affect the steam temperature. The steam temperature deviation must be
kept within 1% of the nominal value. We assume that a feedwater flow increase (FWF) is detected at 4:38:00 and a steam temperature decrease (STT) is detected at 4:40:27. The occurrence time of the steam temperature event is $t_{\text{STT}} = 4:40:27$. The related variables are drum pressure (DRP), steam flow (STF) and spray water flow (SWF). The status of DRP and SWF are “normal” and with positive increase for the case of STF. The SEDRET screen in Fig. 8 shows the relevant information about the steam temperature event.

Fig. 9 shows the conclusions when a steam temperature decrease occurs in the first time interval. If the ST decrease occurs in the first time interval, the initial cause is a power load increase (the steam flow increase) with a probability of 0.92. The recommended actions are (1) check the steam flow and (2) check the spray water flow. The probable past events are a ST flow increase, 10–42 s before the ST temperature decrease, with a probability of 0.96; a drum pressure decrease, 100–272 s before ST temperature decrease, with a probability of 0.62; and spray water flow increase, 42–100 s before the ST temperature decrease, with a probability of 0.36. There is no probable future event because it is a leaf node. Fig. 8 shows the user interface for the analysis of the detection of a decrease in the steam temperature and an increase in the feedwater flow.

SEDERET provides the operators with an earlier and more precise diagnosis events and disturbances. We are encouraged by the fact that the model can produce a reasonable accuracy in times that are compatible with real time decision making. This is important for situations in which the operator must take the best control action to avoid a shutdown of the power plant.

5. Conclusions and future work

This paper presented the design and implementation of an event and disturbance analysis system (SEDERET). SEDRET has three main tasks: construction of the probabilistic network; propagation of the evidences when an event is detected; and explanation of the causes of events and prediction of future events, and their time intervals. This system can be used to assist an operator in real-time assessment of plant disturbances and in this way contribute to the safe and economic operation of power plants. Development of SEDRET was initially motivated by the desire to improve plant availability through early diagnosis and prediction of events and disturbances that could lead to plant shutdown.

An intelligent framework for event and disturbance
analysis has been a primary objective throughout the project. The framework is based on a novel methodology for dealing with uncertainty and time called TNBN. A TNBN generates a formal and systematic structure used to model the temporal evolution of a process under uncertainty. The inference mechanism consists in the propagation of evidences through the net using a probabilistic technique.

TNBN model is based on representing changes of state in each node. If the number of possible state changes for each variable in the temporal range is small, the resulting model is much simpler. This facilitates temporal knowledge acquisition and allows efficient inference using probability propagation. The model supports multiple granularity, with different number of temporal intervals for each node, and different duration for each interval within a node. In contrast, previous probabilistic temporal models are, in general, based on static models repeated at different times. The resulting models are quite complex for many applications and they support a single granularity.

The tests confirmed the ability of SEDRET to properly diagnose disturbances under abnormal situations. Three disturbances were modeled in detail for the steam generator system using data from the training simulator. The primary benefits of the system are: (a) an early and precise diagnosis and prediction of events; and (b) an improvement in the timeliness of the corrective action by the operator. The results of the experiments show the application of SEDRET for power plants. Nevertheless, this formalism can be applied to other dynamic domains, such as clinical diagnosis, planning, and forecasting.

At present, the probabilistic intelligent system is a prototype to be tested in a real fossil power plant. Our future work will be focused on developing and integrating our approach in a real-time diagnostic system and in modeling other industrial domains, such as petroleum refineries.

References


