Proceedings of the International Joint Conference IBERAMIA/SBIA/SBRN 2006 - 3rd Workshop on MSc dissertations and PhD thesis in Artificial Intelligence (WTDIA'2006), Ribeirão Preto, Brazil, October 23–28, 2006. CD-ROM. ISBN 85-87837-11-7

Explanation Generation trough Probabilistic Models for an Intelligent Assistant

Author: Francisco Elizalde[1,3] (PhD. Student)
Supervisor: Enrique Sucar[2] and Co-Supervisor: Pablo deBuen[3]

[1] Tecnológico de Monterrey, Campus Cuernavaca, Paseo de la Reforma 182-A, Col. Lomas de Cuernavaca, Temixco, Morelos, 62589, México
[2] Instituto Nacional de Astrofísica, Óptica y Electrónica, Luis Enrique Erro No. 1, Tonantzintla, Puebla, 72000, México
[3] Instituto de Investigaciones Eléctricas, Reforma 113, Col. Palmira, Cuernavaca, Morelos, 62490, México

Abstract. Under emergency conditions in a complex process, such as a power plant, an operator has to assimilate a great amount of information to promptly analyze the source of the problem, in order to take the corrective actions. To assist the operator to face these situations, we have developed an intelligent assistant system (IAS). An important aspect of the IAS, is its explanation generation mechanism, so that the trainee has a better understanding of the recommended actions and can generalize them to similar situations. In this work, a Markov decision processes (MDP) approach is proposed for explanation generation. For certain training scenario, an optimal policy is obtained by solving an MDP representation of the process. Based on this policy, the IAS advices the trainee on the correct action for given the plant state. When the operator takes an incorrect action, an explanation is generated based on the MDP model. Explanations are predefined by an expert and encapsulated within explanation units. Each explanation unit has an explanation list and a relevant variable that is the most important under the current situation. We have evaluated the IAS with 10 users, half receive an advice with explanations, and the other 5 the advice without explanation. The results show in this first stage that the users with explanations have a better performance. Actually, in a second stage we propose an extension to the explanation mechanism based on a predefined templates to an automatic explanation generation mechanism based on a factorized representation of a MDP.

Keywords: *Intelligent Assistant, Explanation Generation, Probabilistic Models, Markov decision processes.*

1 Research Motivation

Under emergency conditions in a complex process, such as a power plant, an operator has to assimilate a great amount of information to promptly analyze the source of the problem, in order to take the corrective actions. In such situations,

He has to be able to discriminate between erroneous inputs, and to promptly identify the source of the problem in order to define the corrective actions to be taken. To assist the operator to face these situations, we have developed an intelligent assistant system (IAS) to train and assist them [1]. An important requirement for intelligent assistants is to have an explanation generation mechanism, so that the trainee has a better understanding of the recommended actions and can generalize them to similar situations [2].

In this work, as a first stage we extend the IAS by incorporating explanations to it. When an incorrect action is taken by the operator, an explanation is generated based on a MDP model [3]. Adequate explanations are selected according to the current state and the recommended actions are obtained from the MDP. Explanations are predefined by a domain expert and are encapsulated within explanation units. Each explanation unit has an explanation list and a relevant variable that is the most important under the current situation. We have evaluated the IAS with 10 users, half receive the explanations generated, and the other 5 the advice without explanation. We compared the learning gain of both groups, and the results show that the users with explanation have a significantly higher increase in performance, with respect to the users that have only the advice without explanations. The results lead us to establish a proposal as a second stage. This implies to implement an automatic explanation generation mechanism based on a factorized MDP approach and a knowledge base of the domain. The proposal is an extension from the explanation generation mechanism based on a predefined templates.

2 Related work

2.1 Intelligent Assistants

An intelligent assistant system offers several advantages when incorporated to a training simulator. It aids to support on-line decisions, offers off-line training, as well as an explanation and feedback sub-systems. An intelligent assistant system has two sides: the process side and the human side. This implies that an IAS has to understand: (a) the process; (b) the operation model; and (c) the operator's behavior [4]. The process model allows the IAS to anticipate, by simulation, the process behavior, to check the effects of the operator's actions, and to enable the operator to simulate alternative solutions before making a decision. The operation model allows the IAS to control the system. Using a knowledge base, the IAS can eventually correct the operator and suggest alternative sequences. With the operator model, the IAS can infer the operator's intentions and preferences by observing his choices during the problem solving.

Several IAS for plant operators have been developed, such as ASTRAL [5], SOCRATES [6], and SART [7]; however these have very limited explanation capabilities. Previous approaches for operator's training infer the actions and errors of the user from the plant state, and this makes it more difficult to detect the actual errors and give the appropriate explanation feedback. In contrast with them, our system implements an explanation generation approach that directly

evaluates the discrepancies of the user actions with the optimal action and gives a more direct feedback based on explanations.

2.2 Explanation in Probabilistic Graphical Models

Although there is a lot of work for explanation generation for some representations, such as rules and qualitative models, there is very little work on explanation for probabilistic representations, in particular for graphical models. The work on explanations based on graphical models (GM) can be divided according to the classes of models considered, basically Bayesian networks (BN's) and decision networks. BN's [8] graphically represent the dependencies of a set of random variables, and are usually used for estimating the posterior probability of some variables given another. So the main goal of explanations is to try to understand this inference process, and how it propagates through the network. Two main strategies have been proposed for explanation with BN's. One strategy is based on transforming the network to a qualitative representation, and using this more abstract model to explain the relations between variables and the inference process [9], [10]. The other strategy is based on the graphical representation of the model, using visual attributes (such as colors, line widths, etc.) to explain relations between nodes (variables) as well as the the inference process [11]. The explanation of the links represents qualitative influences [12] by coloring the links depending on the kind of influence transmitted from its tail to its head. Another possibility for static explanation consists of explaining the whole network.

Decision networks or influence diagrams extend BN's incorporating decision nodes and utility nodes. The main objective of these models is to help in the decision making process, by obtaining the decisions that maximize the expected utility. So explanation in this case has to do with understanding why some decision (or sequence of decisions) is the optimal one given the current evidence. There is very little work on explanations for decision networks. Bielza [13] proposes an explanation method for medical expert systems based on influence diagrams. It is based on reducing the table of optimal decisions obtained from an influence diagram, building a list that clusters sets of variable instances with the same decision.They propose to use this compact representation of the decision table as a form of explanation, showing the variables that are fixed as a "rule" for certain case. It seems like a very limited form of explanation, difficult to apply to other domains.

MDPs can be seen as an extension of decision networks, that consider a series of decisions in time (dynamic decision network). Thus, the work in explanation for Bayesian and decision nets is relevant, although not directly applicable.

3 Intelligent Assistant

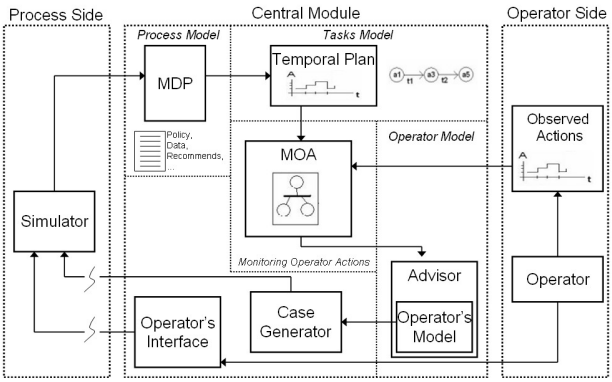

Fig. 1. IAOT description [1].

We have developed an Intelligent Assistant for Operator's Training (IAOT), see Fig. 1. The input to the IAOT is a recommended-plan generated by a decision-theoretic planner, which establishes the sequence of actions that will allow to reach the optimal operation of a steam generator. Operator actions are monitored and discrepancies are detected regarding the operator's expected behavior. This behavior is followed-up using a model (obtained from the optimal plan) represented as a Petri net [14]. IAOT's architectural components are described as follows. Fig. 1 shows the following three main blocks: *process side*, *central module*, *operator side*. The simulator that represents the plant is on the process side. The operator side includes the operator (human-machine interface) and the observed actions. The central module has 5 submodules: *process model*, *task model*, *operator model*, *Monitoring Operator Actions* and *interface*. The process model contains the MDP which generates the optimal policy. In the task model, a temporal plan is generated. The operator's model includes the advisor module. The monitor of the operator's actions contains a Petri net model for the actions sequence. The assistant detects the differences between the executed and recommended actions, gives advice to the trainee and if necessary stops the current case. Depending on the operator's performance, the adviser presents a new case through the case generator module. This module contains several predefined scenarios for selection, with different complexity levels. Through an operator's interface the trainee interacts with the system. This interface represents the objects, the instruments and the application domain information which the operator uses to complete the task.

4 Explanation Generation

4.1 Explanation based on predefined templates

Our proposal for explanation generation trough a probabilistic approach for an intelligent assistant is based on a two stages: in a first stage, adequate actions

are selected based in an action-state for the MDP. In a previous process, such actions were defined by the domain expert and the knowledge were encapsulated in explanation units ($Unid_{Exp}$); in a second stage, an automatic explanation generation mechanism is proposed, since a factorized representation of the MDP, it includes the process variables and their interrelations. This second stage, is supported by the factored representation capabilities for reducing the complexity of the space states and is based on a two state dynamic Bayesian networks [15]

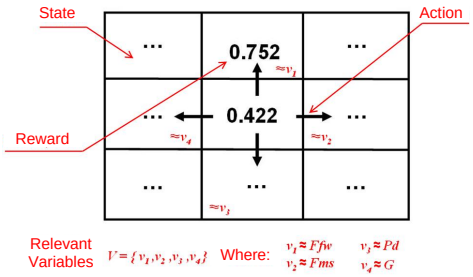

Fig. 2. A simplified representation of the state space, showing the the actions for certain state, including the optimal action and the relevant variables.

The MDP representation establishes the optimal action given a plant state (optimal policy). Explanations are derived from the MDP optimal policy (Fig. 2) by considering: a) the optimal action for the current state; b) the relevant variable (Var_{Rel}) for each action (determined by the expert); and c) the most adequate explanation that justifies the action. These explanations are defined by an expert and are encapsulated in explanation units ($Unid_{Exp}$).

From the optimal policy, the IAOT extracts the two main components to generate an ($Unid_{Exp}$): 1. An optimal action given the state, and; 2. A relevant variable (Var_{Rel}) given the action-state. The first component of the ($Unid_{Exp}$), is an explanation of why the the specific action is selected given the current plan state. In the current implementation, a set of templates were defined for typical plant conditions with the help of an expert in the domain. These templates are stored in a data base. Based on the plant state and the optimal action, the assistant extracts from the data base the most adequate template, which is presented to the user. In the future we plan to generate this templates automatically from a factored representation of the MDP.

The second component is the (Var_{Rel}), that is, the aspect (variable) in the plant that is more critical under the current situation and which is modified by

the correct action. Fig. 2 shows a simplified representation of the state space (considering only 2 variables or dimensions), including the expected value of some of the states obtained by solving the MDP. The arrows illustrate some possible actions. For the state in the middle, the optimal action corresponds to the arrow pointing upwards. In this case, the (Var_{Rel}) is V_1, which is the variable in the Y axis in the state space, modified by the optimal action.

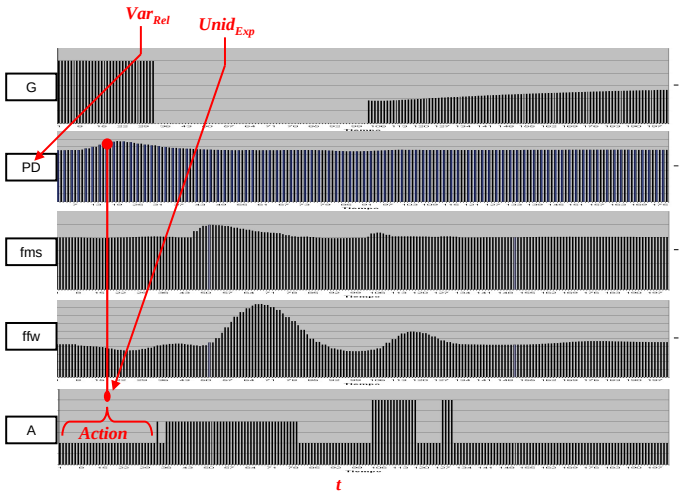

Fig. 3. Temporal evolution of the 4 main variables and the action. The relevant variable, Pd for the state-action at the start of the interval is highlighted.

An example of evolution of some of the variables in our application domain (described in the next section), and the optimal actions are shown in Fig. 3. In this case there are 4 variables, G, Pd, Fms, Ffw. The last graph, A, depicts the actions executed at different times. For the first action, highlighted in the figure, the Var_{Rel} is Pd which has reached a maximum value, and it starts to decrease as a result of this action. The associated $Unid_{Exp}$ will explain this to the operator.

In summary, we have designed an explanation mechanism based on an optimal policy derived from an MDP, and explanation units defined by a domain expert. Using this mechanism, whenever the IAOT detects an error by the user, it gives him advice by displaying the correct actions and the associated explanation. In the next section we show empirical evidence that these explanations help in the learning process.

4.2 Automatic explanations generation

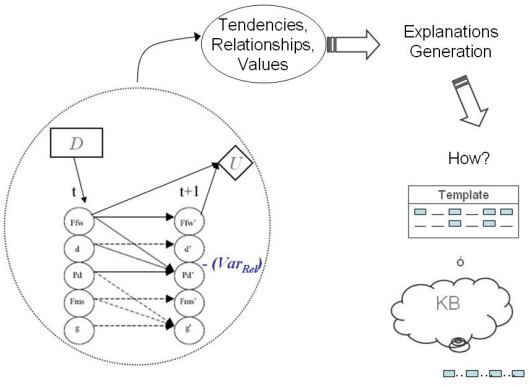

Fig. 4. Proposal for an automatic explanations generation from a MDP factorized representation.

Fig. 4 depicts the proposal approach, since a MDP factorized representation a relevant variable is obtained (Var_{Rel}), such variable is based on decision and utility value (U). With this representation model is possible to obtain values, relationships and tendencies to support an explanations generation mechanism. The structured explanation is based on the fundamental approach when a structure generalizes whatever explanations possibility. It means, that the obtained values from the model fills the empty values in a pre-defined template, and since an expert' knowledge base permits concatenate knowledge units in order to obtained values from the model. The basis of the structured explanation is based on Bielza [13], where, a structured lists call KBM2L (Knowledge Base Matrix to List) proposes an ordered representation of cases that determines the explanation. Each element is represented given a set of ordered indexes with upper and lower limits. Where, limits are the fixed part for the common components of the explanation and the obtained values from the model, are the variable part, for the explanation construction.

Our basic assumption to concatenate knowledge units in order to obtained values from the model is based on a model formulator approach [16]. An inference and a representation method it will be obtained for explanation structuring and for explanations fragments selection. Other options for automatic explanation units concatenation are the qualitative approaches [10], [17], by this option the affected variables into the model are identified, the system behavior and the component functions of the process.

5 Preliminary Results

To evaluate the effect of the explanations on learning, we performed a controlled experiment with 10 potential users with different levels of experience in power plant operation. The set of participants was divided in two groups: 1. Test group (G1): it uses the IAOT with an explanation mode, and has five participants in a three-level profile (novice, intermediate and advanced); 2. Control group (G2): it uses the IAOT without explanations, only advice. Group G2 has a five participants with the same three-level profile.

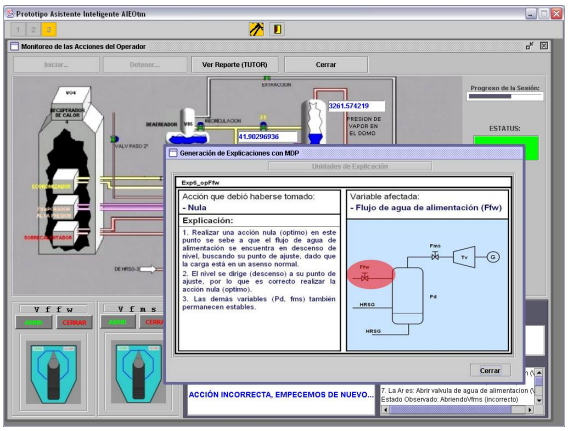

Fig. 5. User interface. In the background it is the simulator, controls and advisor. An explanation is shown in the foreground.

Each participant has to control the plant to reach the optimal state under an emergency condition using a simulator and with the aid of the IAOT. Fig. 5 shows the system interface, including an $Unid_{Exp}$. A record is kept for each session, including if the user reached the goal or how much did he achieved. During each session, the suggested actions and detected errors are given to the user, and for G1, also an explanation.

A case was presented to each user according to his level, and was given up to 5 opportunities to reach the desired state with the help of the IAOT (G1 with explanations, G2 without). Then, without the help of the assistant, the user has to operate the plant (simulator) under a similar situation. Again each participant was given 5 opportunities. For each user in both groups we obtained the percentage of task completion achieved in each opportunity. We adjusted a line to each group to show the tendencies. Fig. 6 summarizes the results, showing a point corresponding to the percentage of task completion for each participant's

opportunity, and a line (obtain with minimum squares) that depicts the general tendency of each group. There is a clear difference between both groups, with a better tendency for the group with explanations.

We consider that these results give evidence that explanations help in the learning of skills such as those required to operate an industrial plant. The hypothesis is that explanations provide a deeper understanding of the process, so the advice can be generalized to similar situations. Of course this needs to be validated with further experiments.

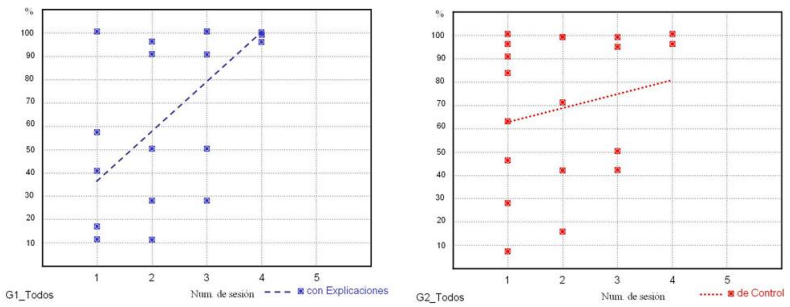

Fig. 6. Graphs comparing the performance of both groups. Left: test group (G1). Right: control group (G2).

6 Conclusions and future work

In this work a new MDP-based approach for explanation generation for an intelligent assistant is presented. From an MDP model of the process, the optimal actions are derived and used within an intelligent assistant for operator's training. When an user error occurs, an explanation is obtained based on the MDP model. Adequate explanations are selected based on the optimal policy. A catalog of explanation units is defined for each domain. Each explanation unit contains an explanations list for the action-state and a relevant variable. An initial evaluation of the effects of explanations in training was performed. The results show a better performance for the users with explanations with respect to those without. The main contributions of this work are: 1. a new generic architecture to explanations generation in an intelligent assistant; 2. an explanations generation mechanism based on MDP's, initially pre-defined and afterwards in an automatic way, and; 3. the experimental validation of the explanation usefulness. To an automatic explanations generation mechanism the use of a factored representation is presented as a proposal.

References

1. Elizalde, F., Sucar, E., deBuen, P.: A prototype of an intelligent assistant for operator's training. In: International Colloquium for the Power Industry, México, CIGRE-D2 (2005)
2. Herrmann, J., Kloth, M., Feldkamp, F.: The role of explanation in an intelligent assistant system. In: Artificial Intelligence in Engineering. Volume 12., Elsevier Science Limited (1998) 107–126
3. Puterman, M.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, New York (1994)
4. Brezillon, P., Cases, E.: Cooperating for assisting intelligently operators. In: Design of Cooperative Systems, INRIA (1995) 370–384
5. Caimi, M., Lanza, C., Ruiz-Ruiz, B.: An assistant for simulator-based training of plant operator. Marie Curie Fellowships Annals **Vol. 1** (1999)
6. Vale, Z., Ramos, C., Silva, A., Faria, L., Santos, J., Fernandez, F., Rosado, C., Marques, A.: Socrates an integrated intelligent system for power system control center operator assistance and training. In: IASTED International Conference on Artificial Intelligence and Soft Computing, Cancun, México (1998) 27–30
7. Brezillon, P., Naveiro, R., Cavalcanti, M., Pomerol, J.: Sart, an intelligent assistant system for subway control. Pesquisa Operacional, Brazilian Operations Research Society **20(2)** (2000) 247–268
8. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA (1988)
9. Druzdzel, M.: Explanation in probabilistic systems: is it feasible? is it work? In: Intelligent information systems V, Proceedings of the workshop, Poland (1991) 12–24
10. Renooij, S., Van-DerGaa, L.: Decision making in qualitative influence diagrams. In: Proceedings of the Eleventh International FLAIRS Conference, Menlo Park, California, AAAI Press (1998) 410–414
11. Lacave, C., Atienza, R., Diez, F.: Graphical explanations in bayesian networks. In: Lecture Notes in Computer Science. Volume 1933., Springer-Veralg (2000) 122–129
12. Wellman, M.: Graphical inference in qualitative probabilistic networks. Networks: an international journal **20** (1990) 687–701
13. Bielza, C., del Pozo, J.F., Lucas, P.: Optimal decision explanation by extracting regularity patterns. In Coenen, F., Preece, A., Macintosh, A., eds.: Research and Development in Intelligent Systems XX, Springer-Verlag (2003) 283–294
14. Petri, C.: Kommunikation mit Automaten. Phd. thesis, Faculty of Mathematics and Physics at the Technische Universitt, Darmstadt, Germany (1962)
15. Dearden, R., Boutillier, R.: Abstraction and approximate decision-theoretic planning. Artificial Intelligence **89** (1997) 219–283
16. Biris, E., Shen, Q.: Automatic modelling using bayesian networks for explanation generation. In: Qualitative Reasoning 1999, Loch Awe, Scotland (1999)
17. Bolt, J., Gaag, L.V.D., Renooij, S.: Introducing situational influences in qpns. In Nielsen, T., Zhang, N., eds.: Proceedings of the Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Springer-Verlag (2003) 113 – 124