Explanation Generation trough Probabilistic Models for an Intelligent Assistant

Francisco Elizalde (Std.)[1,2]
[1]ITESM, C. Cuernavaca
Reforma 182-A, Temixco,
Morelos, 62589.
fef@iie.org.mx

Enrique Sucar (Advisor)
INAOE, L. Enrique Erro No.
1, Tonantzintla, Puebla,
72000.
esucar@inaoep.mx

Pablo De Buen (Advisor)
[2]IIE, Reforma 113, Col.
Palmira, Cuernavaca,
Morelos, 62490.
debuen@iie.org.mx

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. 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 [5]. 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 [7].

2. Related work

An IAS supports on-line decisions, offers off-line training, as well as an explanation and feedback sub-systems [2]. Several IAS for plant operators have been developed, such as ASTRAL [3], SOCRATES [9], and SART [2]. However these have very limited explanation capabilities. Explanations based on probabilistic representations can be divided into Bayesian networks (BN's) and decision networks (DN's). One strategy for BN's is based on transforming the network to a qualitative representation to explain the relations between variables and the inference process [4], [8]. The other strategy is based on the graphical representation of the model [7]. DN's extend BN's incorporating decision nodes and utility nodes. The main objective is to help in decision making by obtaining decisions that maximize the expected utility. In DN's for explanations, Bielza [1] proposes an explanation method reducing the table of optimal decisions obtained from DN's, building a list that clusters sets of variable instances with the same decision. MDPs can be seen as an extension of DN's that consider a series of decisions in time. Thus, the work in explanation for BN's and DN's is relevant, but not directly applicable.

2. Proposal

The main questions of this research are: Is it possible to generate explanations based on probabilistic methods? And, if it so, do explanations based on a probabilistic mechanisms improve the operator's performance? Our proposal is based on two stages. At a first stage, adequate explanations are selected based on an action-state of an MDP. In a previous process, such explanations were defined by the domain expert and the knowledge were encapsulated within explain units. At a second stage, an automatic explanation generation mechanism is proposed, based on a factorized representation of the MDP. This second stage can derived a factorized representation of the MDP based on a two state dynamic Bayesian network. Explanations are derived from the optimal policy by considering: a) the optimal action for the current state; b) the relevant variable for each action; and c) the most adequate explanation that justifies the action. Initially, explanations are defined by an expert. From the optimal policy, the IAS (Fig. 1) extracts the two main components to generate an explanation unit: 1. An optimal action given the state, and; 2. A relevant variable given the action-state.

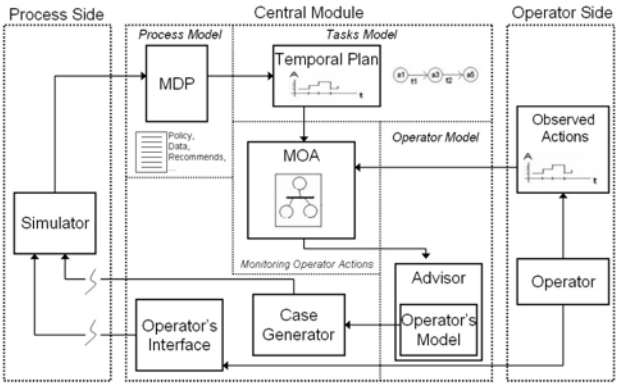

Fig. 1. IAS block diagram [5]

In a second phase, the explanation will be generated automatically from the MDP represented in a factored

form. For this, we will design an explanation "template" based on the expert's explanations units. The slots in the explanation template will include the relevant variables and other factors related to the state-action in the MDP. These slots will be filled based on a qualitative representation of the factored MDP.

3. Preliminary results

We performed a controlled experiment with 10 potential users in two groups: (G1) uses the IAS with an explanation mode, and has five participants in a three-level profile; (G2) uses the IAS without explanations, only advice. During each session, the suggested actions and detected errors are given to the user, and for G1, also an explanation. Fig. 2 summarizes the results, showing a point corresponding to the percentage of task completion for each participant's opportunity and a line (obtained with minimum squares) that depicts the general tendency of the group. There is a clear difference between both groups, with a better tendency for the group with explanations. The hypothesis is that explanations provide a deeper understanding of the process.

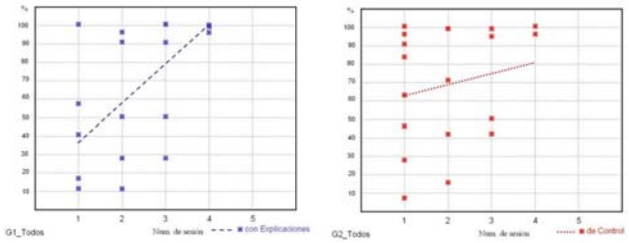

Fig. 2. Graphs comparing the performance of G1 and G2.

4. Conclusions and contributions

A new MDP approach for explanation generation in 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 training. When an error occurs, an explanation is obtained based on the MDP model. 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 for explanations generation in an IAS; 2) an explanations generation mechanism based on MDP's, initially pre-defined and afterwards generated automatically, and; 3) the experimental validation of the explanation usefulness.

10. References

[1] C. Bielza, J.A. Fernández del Pozo and P. Lucas, "Optimal Decision Explanation by Extracting Regularity Patterns", RDIS XX, Springer-Verlag. 2003. pp. 283-294

[2] P. Brezillon, R. Naveiro, M. Cavalcanti and J.Ch. Pomerol, "SART, an intelligent assistant system for subway control", Pesquisa Operacional, BORS. 2000, (20-2) pp. 247-268

[3] M. Caimi, C. Lanza and B. Ruiz-Ruiz, "An Assistant for Simulator-Based Training of Plant Operator", MCFA. Vol. 1, 1999

[4] M.J. Druzdzel, "Explanation in probabilistic systems: is it feasible? is it work?", Intelligent information systems V, Proceedings of the workshop. Poland, 1991, pp. 12-24

[5] F. Elizalde, E. Sucar and P. deBuen, "A prototype of an intelligent assistant for operator's training", CIGRE-D2, International Colloquium for the Power Industry, México, 2005.

[6] J. Herrmann, M. Kloth and F. Feldkamp, "The role of explanation in an intelligent assistant system", Artificial Intelligence in Engineering, Elsevier Science Limited, 1998, pp. (12) 107-126

[7] C. Lacave, R. Atienza and F.J. Diez, "Graphical explanations in Bayesian networks", Lecture Notes in Computer Science, Springer-Veralg , 2000. (1933)122-129

[8] S. Renooij and L. Van-DerGaa, "Decision Making in Qualitative Influence Diagrams", Proceedings of the Eleventh International FLAIRS Conference, AAAI Press. Cal. USA. 1998, pp. 410-414

[9] Z.A. Vale, C. Ramos, A. Silva, L. Faria, J. Santos, F. Fernandez, C. Rosado and A. Marques, "SOCRATES an integrated intelligent system for power system control center operator assistance and training", IASTED on AI and SC. Mexico, 1998. pp. 27-30