Policy Explanation in Factored Markov Decision Processes

Authors:
Francisco Elizalde, ITESM & IIE, Mexico
L. Enrique Sucar INAOE, Mexico
Manuel Luque and Javier Díez, UNED, Spain
Alberto Reyes, IIE, Mexico

September 17, 2008

Overview

1) Motivation
2) Intelligent assistant
3) Automatic explanations
4) Relevant variable selection
5) Experiments
6) Conclusions
7) Current and Future work

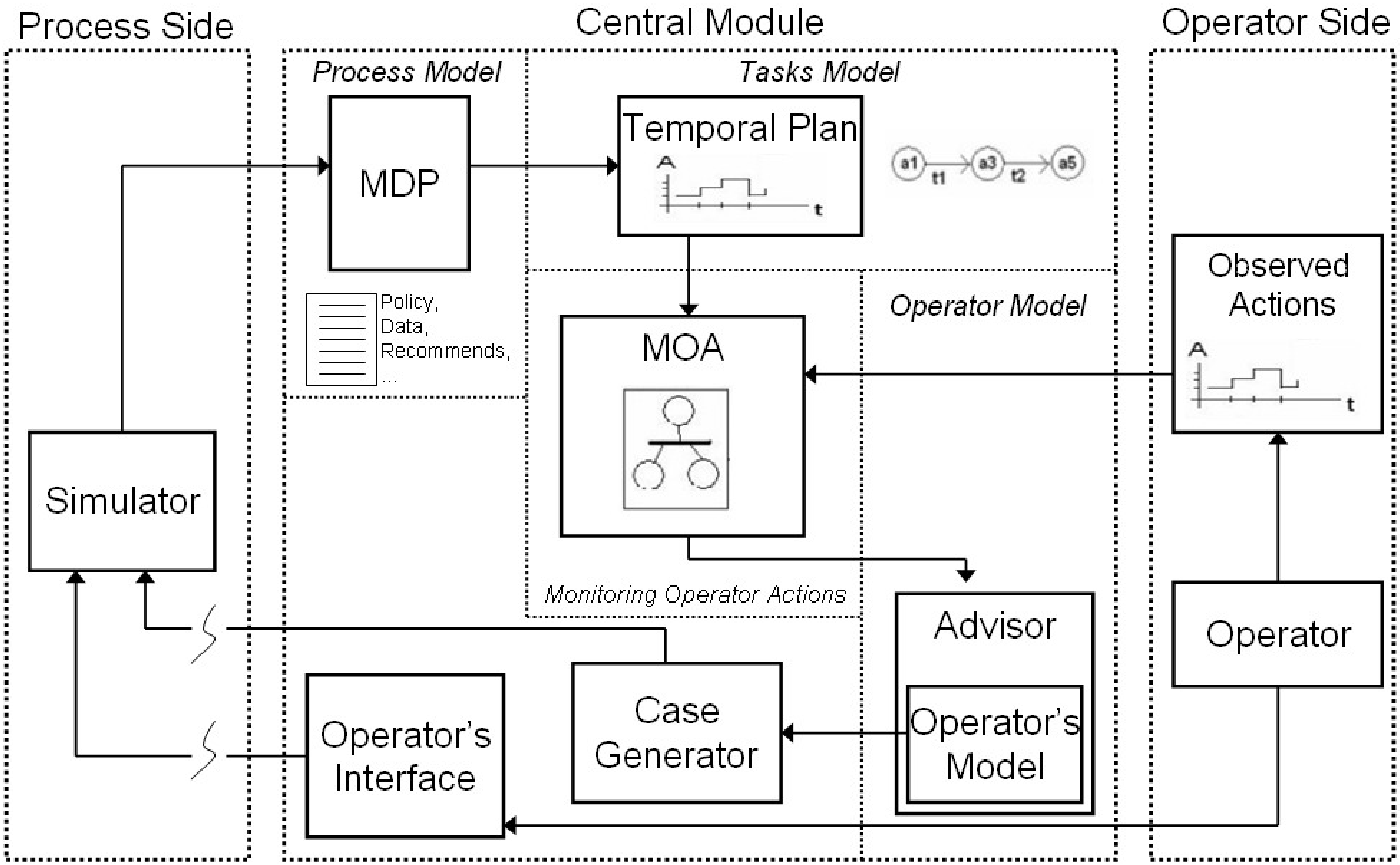
Research motivation

- Under emergency conditions in 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.

Research motivation

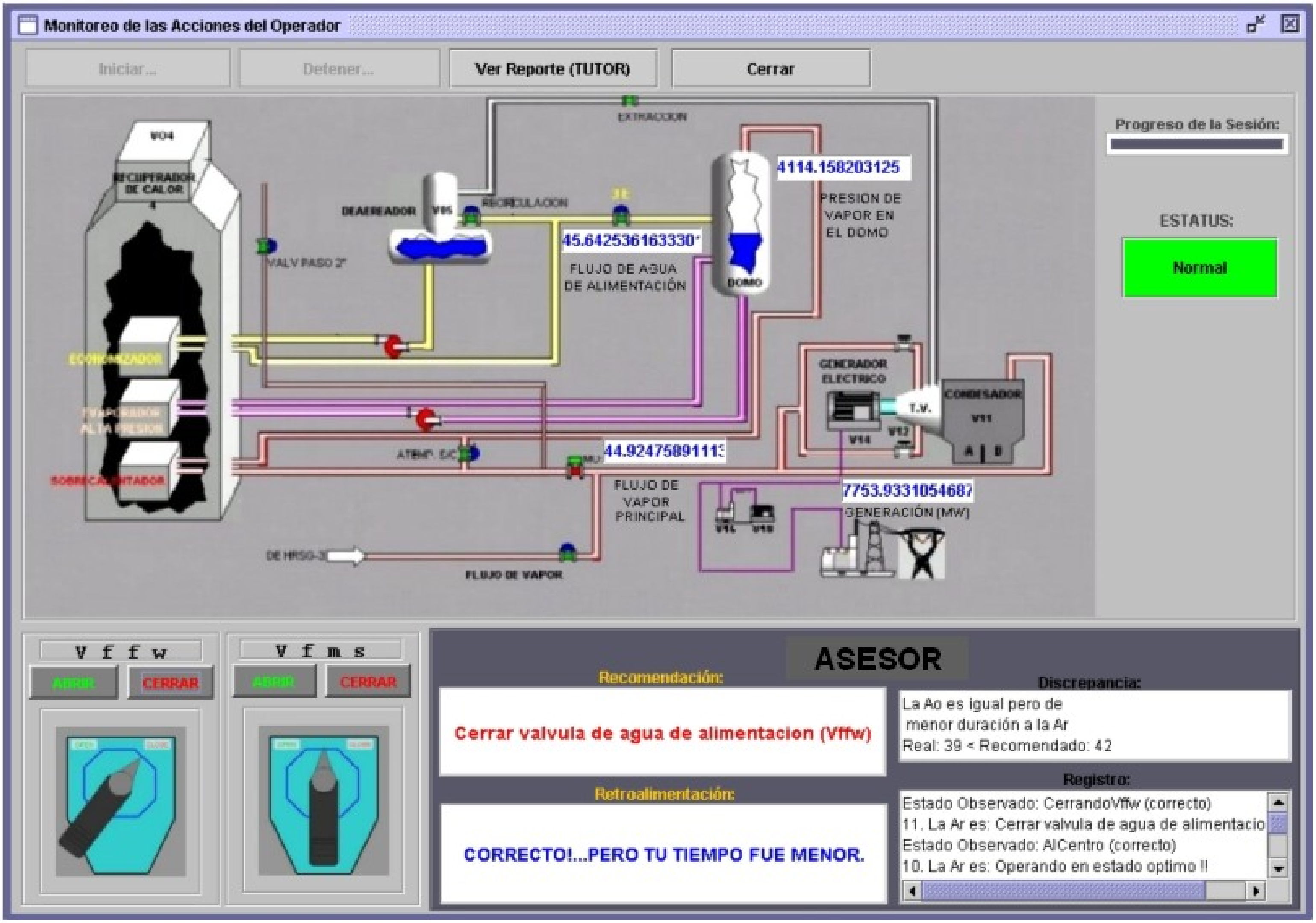
- To assist the operator to face these situations, we developed an intelligent assistant system (IAS) for training and providing recommendations on-line
- The recommendation are based on an MDP that has been previously solved to obtain the optimal policy: the action that the operator should do in each situation
- 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.

IAS Architecture



Intelligent Assistant for Operator's Training (IAOT) [Elizalde et Al., 2005]

User Interface



4114.158203125

45.642536163330

44.92475891113

7753.9331054687

Recomendación:

Cerrar valvula de agua de alimentacion (Vffw)

Retroalimentación:

CORRECTO!...PERO TU TIEMPO FUE MENOR.

Discrepancia:

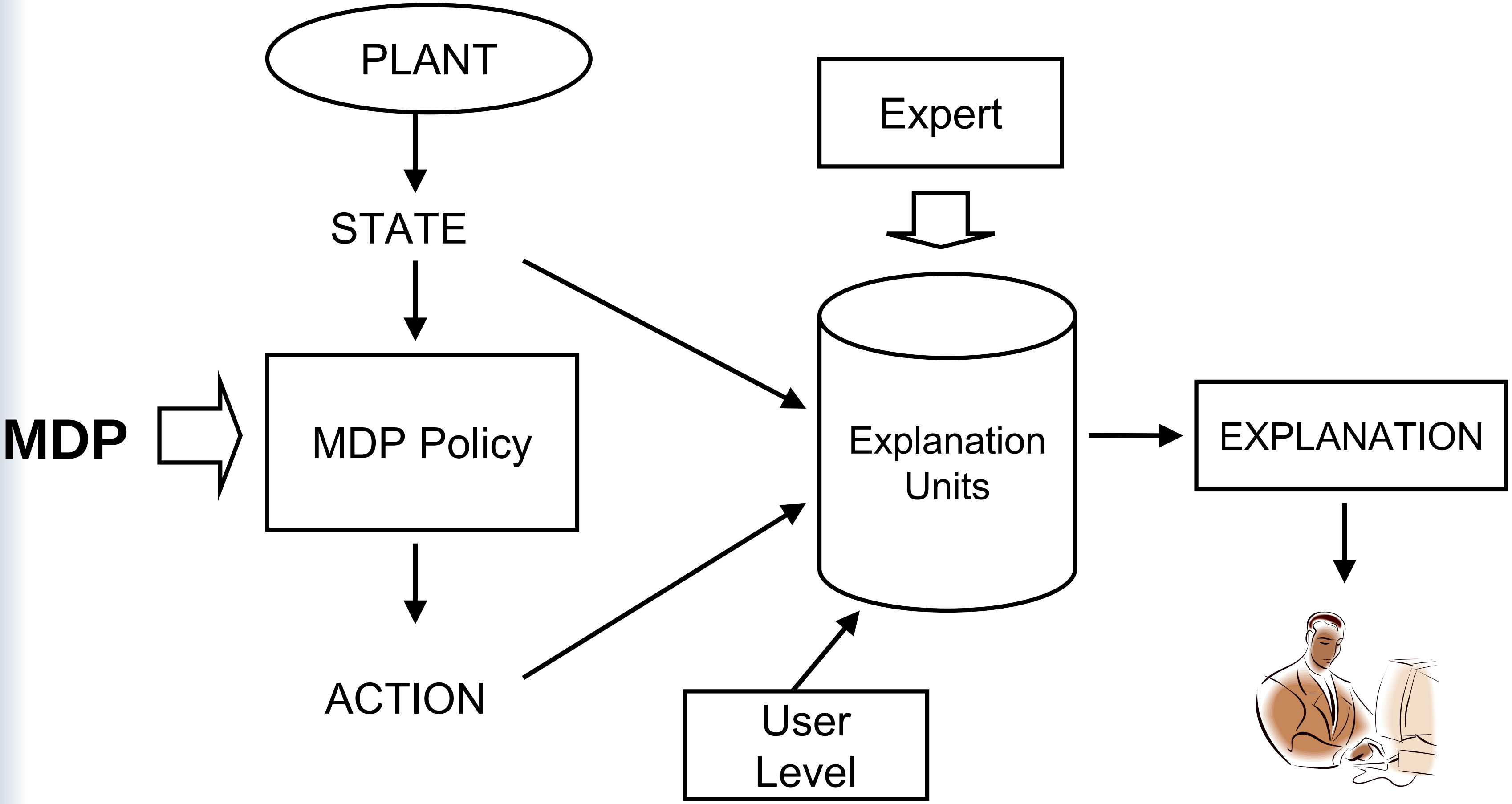
La Ao es igual pero de
menor duración a la Ar
Real: 39 < Recomendado: 42

Registro:

Estado Observado: CerrandoVffw (correcto)
11. La Ar es: Cerrar valvula de agua de alimentacio
Estado Observado: AlCentro (correcto)
10. La Ar es: Operando en estado optimo !!

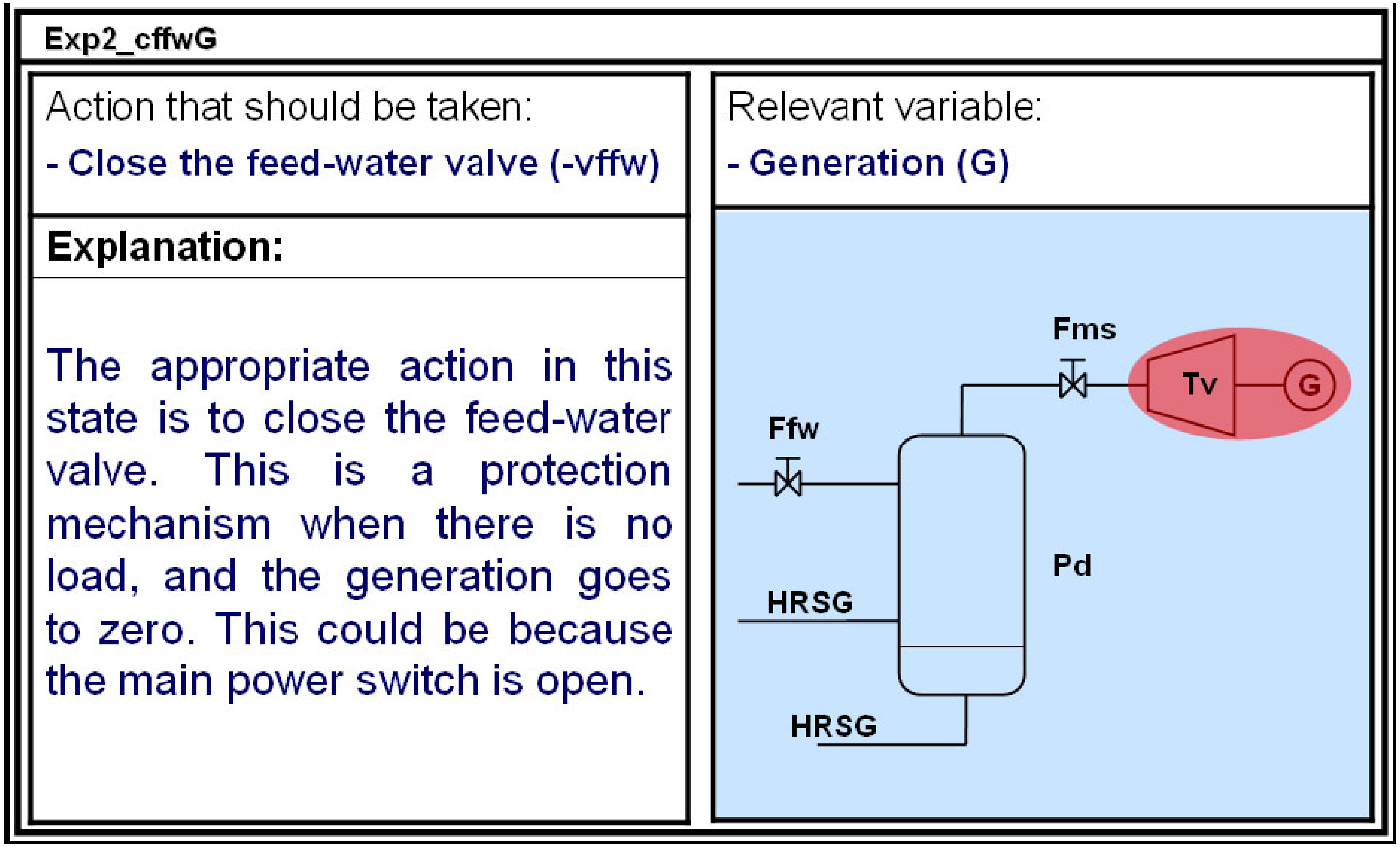
Intelligent Assistant for Operator’s Training (IAOT) [Elizalde et Al., 2005]

Built-in explanations



Explanation generation from a MDP

Example of an explanation unit



Exp2_cffwG

Action that should be taken:

- Close the feed-water valve (-vffw)

Relevant variable:

- Generation (G)

Explanation:

The appropriate action in this state is to close the feed-water valve. This is a protection mechanism when there is no load, and the generation goes to zero. This could be because the main power switch is open.

Experiments: first stage

- To evaluate the impact of explanation on learning we conducted a controlled user study.
- Several potential users solved different cases using a power plant simulator.
- They received advice from the IAS: some with explanations and some without.
- We compared both groups in terms of the number of trails required to reach the goal without errors

Conclusions from the first-stage

-The results of this experiments show a significant difference in favor of the group that had explanations.

-Since obtaining the explanations from an expert is a complex and time-consuming process, it is desirable that the assistant can generate the explanations automatically.

Automatic Explanation Generation

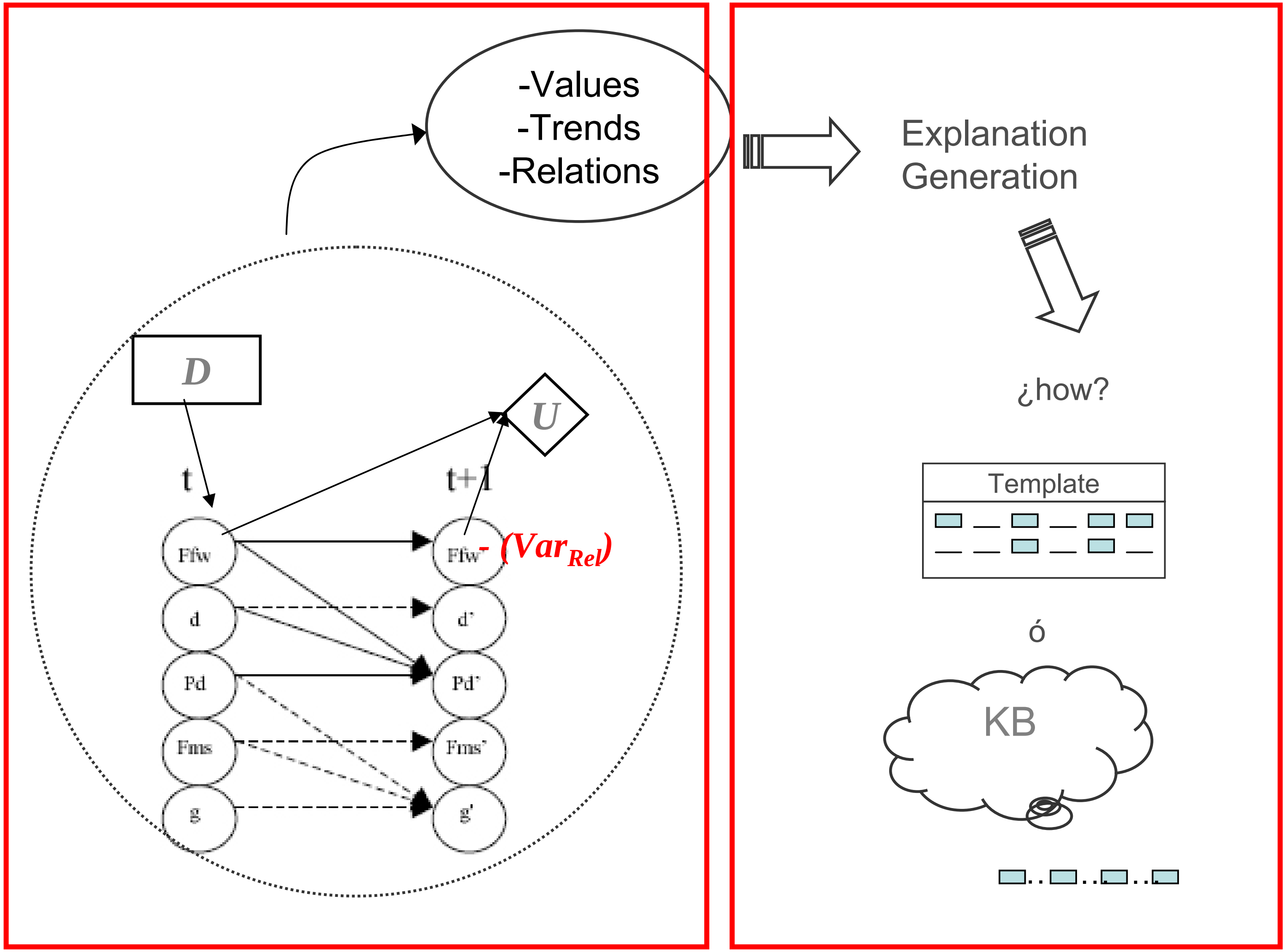
- Analyzing the explanations provided by the expert, we discovered that in all the cases, the explanation starts for a ***relevant variable***; the variable that is most important under the current situation.
- So our strategy for explanation generation starts by finding this relevant variable for each state.

Automatic Explanation Generation

- The basic idea is to consider the factored representation of the MDP, where the transition function is represented as a two-stage dynamic Bayesian network.

- Based on factored MDP representation, we want to determine which of the variables is the most “important” for a given state-action.

Automatic explanation generation



-Values
-Trends
-Relations

Explanation Generation

¿how?

Template

ó

KB

Factored Model

Explanations

Relevant variable selection

We propose two heuristic rules for obtaining the relevant variable, one based on utility and other based on policy:

Utility-based

The utility–based rule evaluates how much the utility function will change if we vary the value of one of the variables for the current state, keeping the other variables fixed.

Policy-based

The policy–based rule estimates the potential changes in optimal action for each of the variables.

Utility-based rule

Let us assume that the process is in state s, then we measure the *relevance of a variable* $\boldsymbol{X_i}$ *for the state* $\boldsymbol{s}$ *based on utility*, denoted by $rel_s^V\ (X_i)$, as:

$$rel_s^V(X_i) = \max_{s' \in neigh_{X_i}(s)} V(s') - \min_{s' \in neigh_{X_i}(s)} V(s')$$

- Where $neigh_{Xi}(s)$ is the set of states that take the same values than s for all other variables X_j, $j \neq i$; and a different value for the variable of interest, X_i.

- That is, the maximum change in utility when varying the value of X_i with respect to its value under the current state s.

Utility-based rule

This expression is evaluated for all the variables, and the one with the highest value is considered the most relevant for state ***s***, according to the value criteria:

$$X_R^V = argmax_i(rel_s^V(X_i)), \forall(i)$$

Policy-based rule

- The second heuristic rule for finding the most relevant variable consists in exploring the optimal policy to detect changes in the optimal action for the state.
- The variable that may cause more changes in policy will be selected as the most relevant.

Policy-based rule

Let us assume that the MDP is in state ***s***, then we measure the *relevance of a variable X_i for the state **s** according to its impact on policy*, denoted by $rel_s^A(X_i)$, as:

$$rel_s^A(X_i) = \#s' : s' \in neigh_{X_i}(s) \wedge \pi^*(s) \neq \pi^*(s')$$

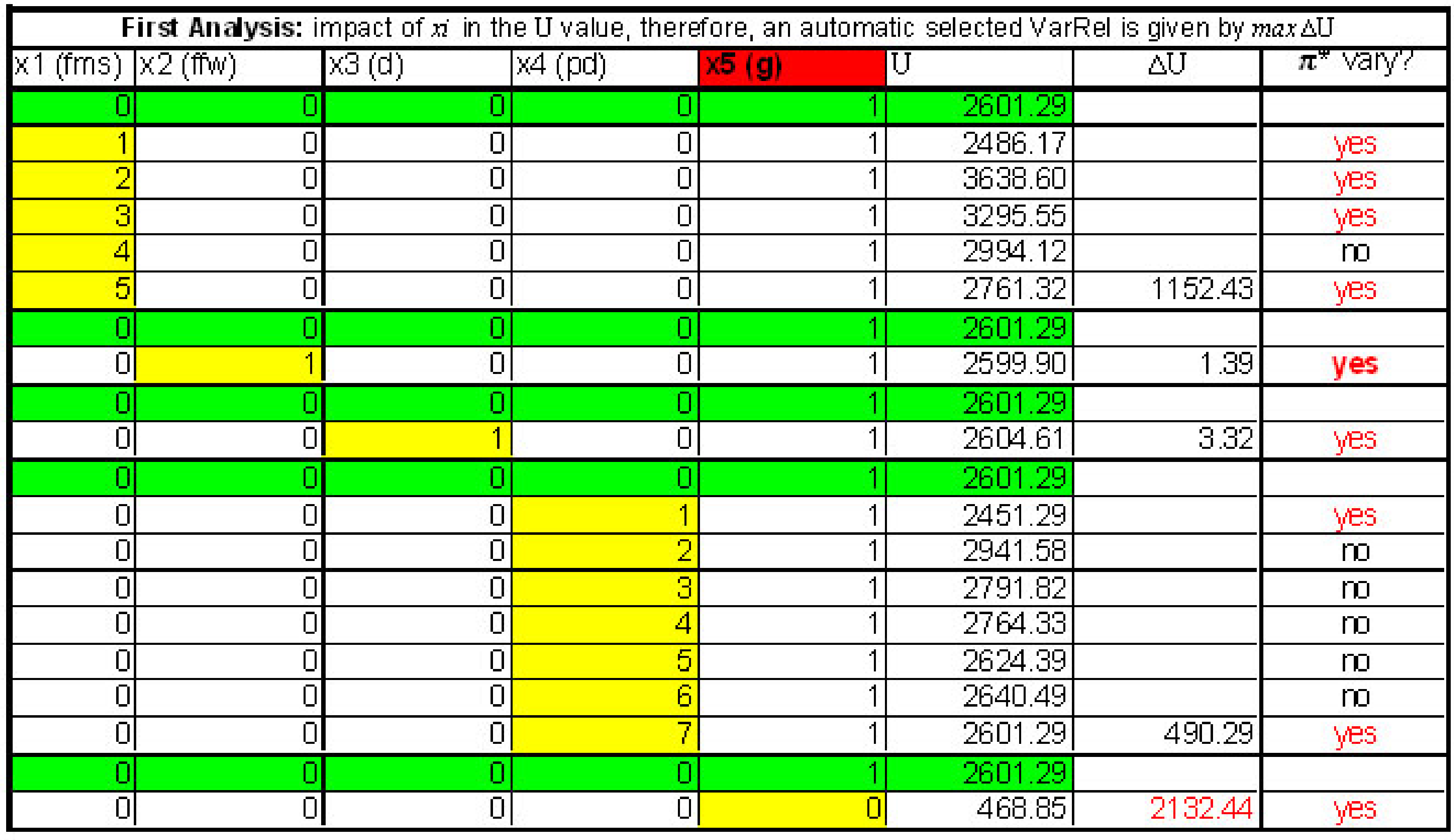
- Where $neigh_{Xi}(s)$ is the set of states that take the same values than ***s*** in all the variables except in variable X_j, *π*(s)* is the optimal action under the current state s, and *π*(s')* is the action that will be taken in the other states such that s' Є $neigh_{Xi}(s)$.

Policy-based rule

This expression is evaluated for all the variables, and the one with the highest value is considered the most relevant for state ***s***, according to the value criteria:

$$X_R^A = argmax_i(rel_s^A(X_i)), \forall(i)$$

Example

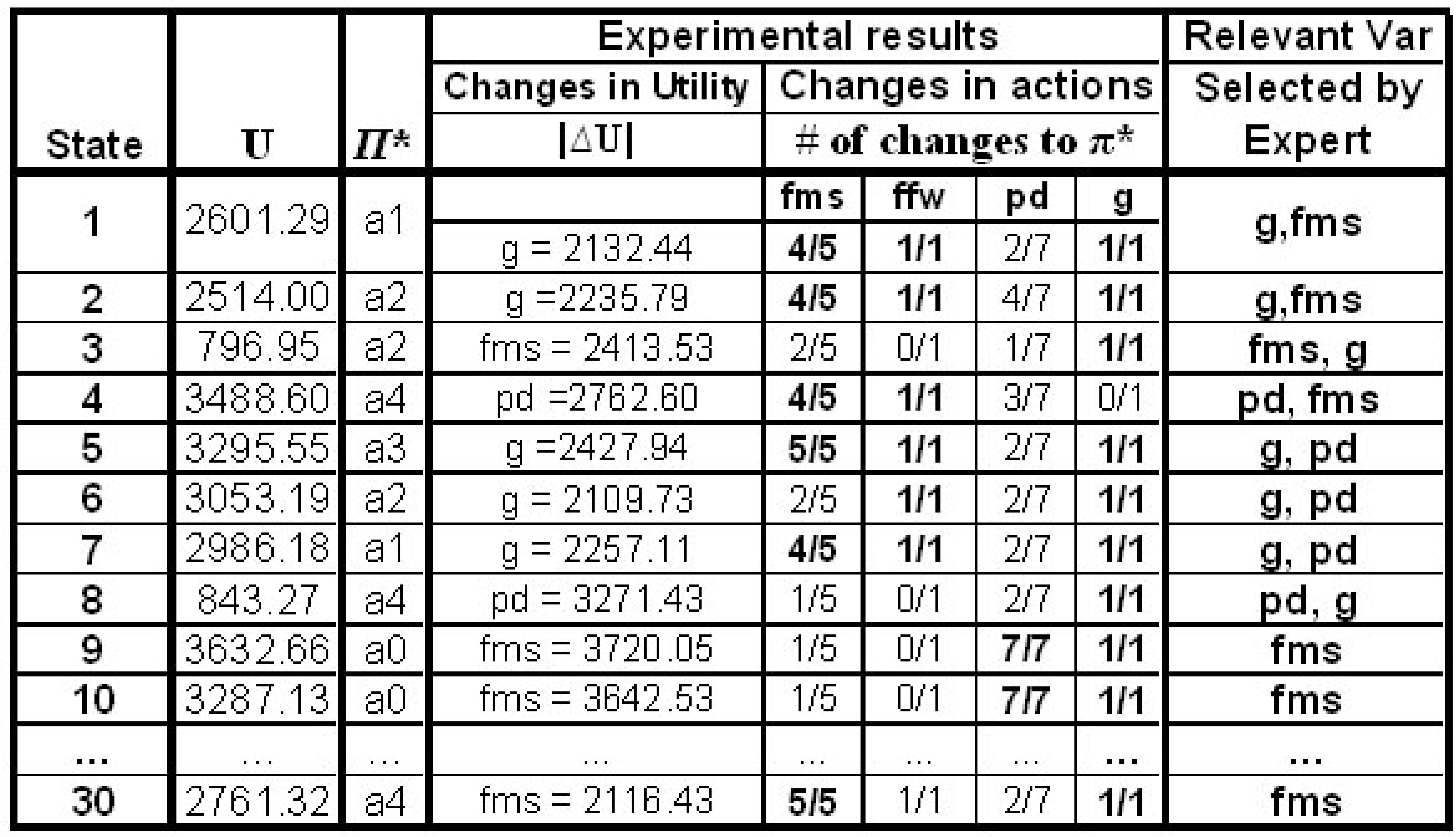


First Analysis: impact of x_i in the U value, therefore, an automatic selected VarRel is given by $max\,\Delta U$							
x1 (fms)	x2 (ffw)	x3 (d)	x4 (pd)	x5 (g)	U	ΔU	π* vary?
0	0	0	0	1	2601.29		
1	0	0	0	1	2486.17		yes
2	0	0	0	1	3638.60		yes
3	0	0	0	1	3295.55		yes
4	0	0	0	1	2994.12		no
5	0	0	0	1	2761.32	1152.43	yes
0	0	0	0	1	2601.29		
0	1	0	0	1	2599.90	1.39	**yes**
0	0	0	0	1	2601.29		
0	0	1	0	1	2604.61	3.32	yes
0	0	0	0	1	2601.29		
0	0	0	1	1	2451.29		yes
0	0	0	2	1	2941.58		no
0	0	0	3	1	2791.82		no
0	0	0	4	1	2764.33		no
0	0	0	5	1	2624.39		no
0	0	0	6	1	2640.49		no
0	0	0	7	1	2601.29	490.29	yes
0	0	0	0	1	2601.29		
0	0	0	0	0	468.85	2132.44	yes

Experiments

We compared the relevant variables obtained with these rules with the one given by the expert, for a representative sample of states (30) in the power plant domain.

In general there was a strong agreement, which contributes evidence to the validity of the proposed approach.



State	U	*Π**	Experimental results					Relevant Var
			Changes in Utility	Changes in actions				Selected by
			|ΔU|	# of changes to π*				Expert
				fms	ffw	pd	g	
1	2601.29	a1	g = 2132.44	**4/5**	**1/1**	2/7	**1/1**	g,fms
2	2514.00	a2	g =2235.79	**4/5**	**1/1**	4/7	**1/1**	g,fms
3	796.95	a2	fms = 2413.53	2/5	0/1	1/7	**1/1**	fms, g
4	3488.60	a4	pd =2762.60	**4/5**	**1/1**	3/7	0/1	pd, fms
5	3295.55	a3	g =2427.94	**5/5**	**1/1**	2/7	**1/1**	g, pd
6	3053.19	a2	g = 2109.73	2/5	**1/1**	2/7	**1/1**	g, pd
7	2986.18	a1	g = 2257.11	**4/5**	**1/1**	2/7	**1/1**	g, pd
8	843.27	a4	pd = 3271.43	1/5	0/1	2/7	**1/1**	pd, g
9	3632.66	a0	fms = 3720.05	1/5	0/1	**7/7**	**1/1**	fms
10	3287.13	a0	fms = 3642.53	1/5	0/1	**7/7**	**1/1**	fms
...	...	...	...	...	...	...	...	...
30	2761.32	a4	fms = 2116.43	**5/5**	1/1	2/7	**1/1**	fms

Conclusions

•We developed a method for determining the most relevant variable for generating explanations based on a factored MDP.

•The explanations provided by human experts are based on what they consider the most relevant variable in the current state, so obtaining this variable is an important first stage for automatic explanation generation.

• For determining the most relevant variable we proposed and compared two heuristic rules:

- One based on the impact on utility of each variable,
- and other based on their impact on the policy.

Conclusions

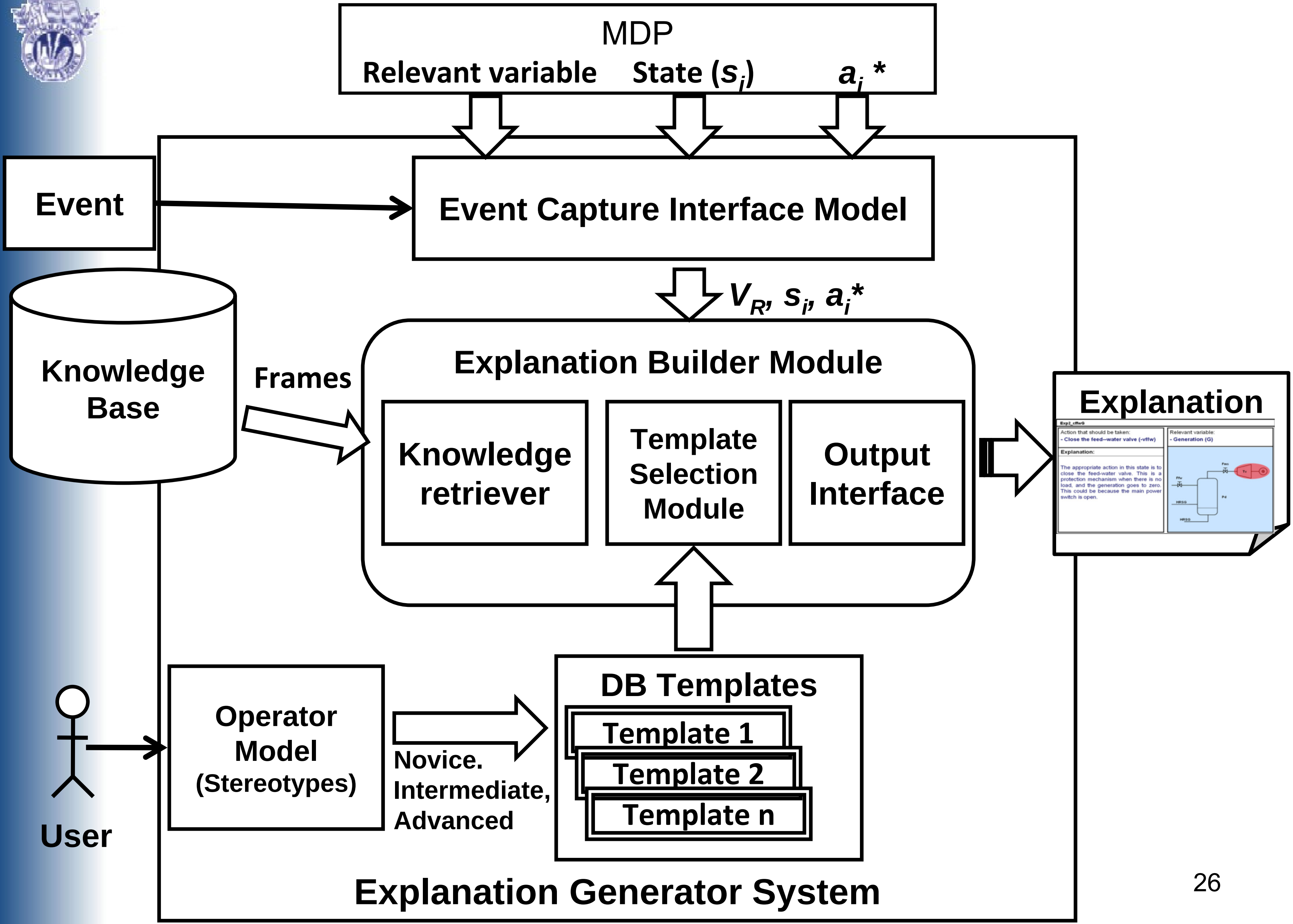
- The experimental evaluation shows that the methodology is promising, as the relevant variables selected agreed, in general, with those chosen by the expert.

- The rule based on utility impact seems more appropriate, at least in this domain, as it gives more specific results with a very high accuracy.

Current and Future Work

- Based on the relevant variable and domain knowledge (represented as frames), we are developing an explanation generator that uses templates.

- The explanation generator uses the current state and optimal action (from the MDP), and the relevant variable, to extract the information from the frame system and fill in the templates [ECAI 2008].



MDP

Relevant variable State (s_i) a_i *

Event

Event Capture Interface Model

V_R, s_i, a_i*

Knowledge Base

Frames

Explanation Builder Module

Knowledge retriever

Template Selection Module

Output Interface

Explanation

Exp2_cffwG

Action that should be taken:

- Close the feed--water valve (-vffw)

Relevant variable:

- Generation (G)

Explanation:

The appropriate action in this state is to close the feed-water valve. This is a protection mechanism when there is no load, and the generation goes to zero. This could be because the main power switch is open.

User

Operator Model (Stereotypes)

Novice. Intermediate, Advanced

DB Templates

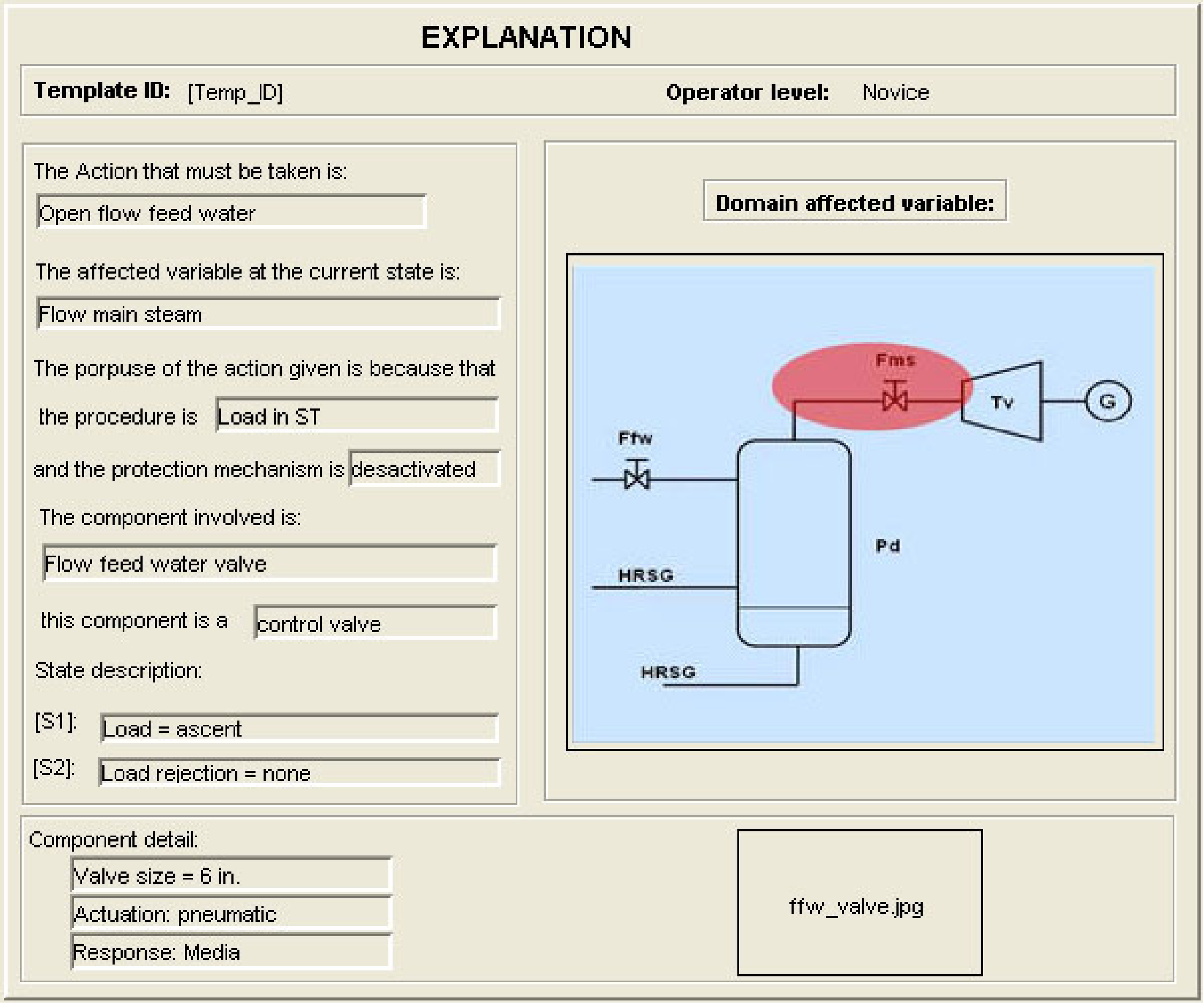
Template 1

Template 2

Template n

Explanation Generator System

Explanation unit



EXPLANATION

Template ID: [Temp_ID] **Operator level:** Novice

The Action that must be taken is: Open flow feed water

The affected variable at the current state is: Flow main steam

The porpuse of the action given is because that the procedure is Load in ST and the protection mechanism is desactivated

The component involved is: Flow feed water valve

this component is a control valve

State description:

[S1]: Load = ascent

[S2]: Load rejection = none

Domain affected variable:

Component detail:

Valve size = 6 in.

Actuation: pneumatic

Response: Media

ffw_valve.jpg

Current and Future Work

- We are currently evaluating the generated templates by comparing them with the expert's explanations.
- In the future we plan to conduct further tests for more cases, and applied this mechanism to other domains.

Thank you!

Questions?