To be presented (and included in the Proceedings) at the 13th International Workshop on Qualitative Reasoning (QR '99).

Automatic Modelling using Bayesian Networks for Explanation Generation

Elias Biris
School of Artificial Intelligence, The University of Edinburgh,
80 South Bridge, Edinburgh EH1 1HN, UK.
Email: eliasb@dai.ed.ac.uk

This work was jointly done with Q. Shen. It was partly supported by an EU grant (BRMA-CT96-5002). The author is grateful to Chris Mellish, James Kwaan and Roderick McKinnel for helpful discussions.

Abstract

The task of generating informative explanations in industrial training involves automated formulation of system models with respect to the varying levels of the trainees' knowledge. Compositional Modeling provides a useful basis upon which to structure a suite of models that may reflect different complexities of the system being modelled. However, additional inferences are required in order to select appropriate model fragments to form a coherent system model that is suitable for a given trainee's degree of expertise. This paper presents a novel approach to perform such inferences by the use of Bayesian networks. The work is implemented and typical experimental results are given.

1 Introduction

The need for informative explanations regarding the behaviour of physical systems arises in many tasks in science and engineering. In industrial training, such explanations are especially significant for the establishment of coherent and consistent knowledge of the components and their associated processes of a given plant. The task of explanation generation involves, essentially, finding *information* that is relevant to a *communicative goal* set by the explainee, from available knowledge sources, and organising this knowledge into a cohesive and coherent multi-sentential text. An important requirement of generating such explanations is the ability to vary the explanation content according to the expertise of the explainee, by adjusting the level of detail of the underlying domain knowledge.

To achieve the required adjustments of domain knowledge, a technique that allows for systematic variation of the knowledge representation is needed. Compositional Modelling (CM) [Falkenheiner and Forbus, 1991], [Gruber and Gautier, 1993], [Nayak, 1994], [Levy *et al.*, 1997] has been developed as a methodology for formulating knowledge models for the domain of interest by *composing* model fragments, i.e. (partial) models of the domain's primitive elements that describe only some aspects of the components' behaviour. As such, CM enables the variation of detail of the entire model by altering the detail of the fragments used as building blocks and is used as basis for the present work.

By reflecting the user expertise to the detail level of at least some of the model fragments, guidelines for the selection of the remaining fragments can be set in order to formulate a model that corresponds to the understanding ability of the user. This is, however, a flexibility that has not been provided in the existing work for automatic model formulation. An approach is herein presented towards enabling such model formulation for the domain objects of interest. The selection of the appropriate model fragments is based on the utilisation of reasoning with a Bayesian network. This is motivated by the intention to employ an efficient as well as formal theory to handle the uncertainty involved with the selection of fragments, based on initial preference of some of the available model fragments according to the information request of the user.

The rest of this document is organised as follows. Section 2 provides an overview of the design of the entire computer program that performs explanation generation, putting the model formulation module in context with the rest of the modules involved. In Section 3, after a brief review of the CM paradigm and basic ideas in reasoning with Bayesian networks, the proposal for model formulation through Bayesian model fragment selection is presented. A simple example of the reasoning involved in this approach is illustrated in Section 4. Section 5 concludes the paper.

2 The Explanation Generation System

The present work is developed within the general framework of explanation generation. Figure 1 shows the main functional modules (Explanation Generator, Approximate Reasoner and Model Formulator) and their component dependencies of an explanation system, designed and implemented by the author.

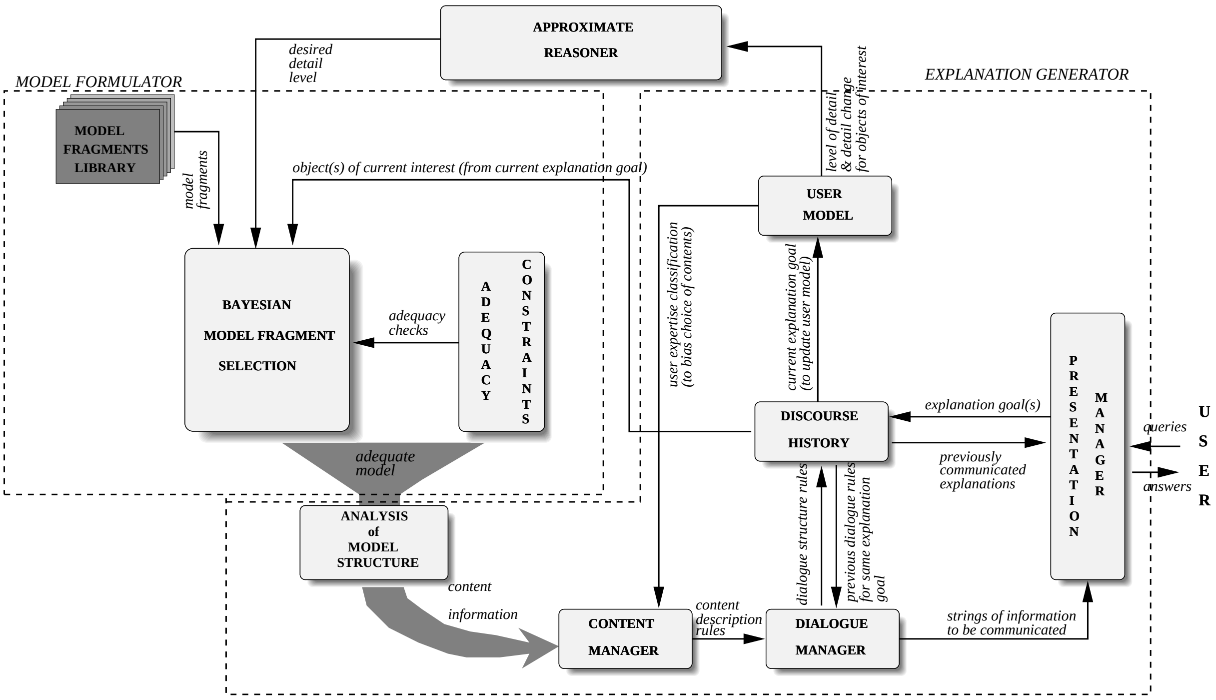


Figure 1: *Explanation system*: individual modules and their components

The Explanation Generator serves as a mediator with the user. During a training session, the user can select questions from predefined menus about the components of the domain system that is presented to the user. Each selected question is internally transformed into a first-order predicate structure which forms the current *user explanation goal*: `question_type{< list of related objects >}`, where an object is a domain component or process of interest. This goal is utilised to update the user model and the history of the discourse, following the methods given in [Cawsey, 1993].

The user model contains a set of profiles for the user, classified according to the object(s) appearing in each of the queries. These profiles essentially keep a record of the object's detail level as it was represented in the model used for the most recent explanation involving this object. The initial detail level of an object is set by default to the simplest level possible for that object. The detail level is determined, currently, by the number of variables involved in the set of equations constituting the model fragment that represents the object. From these profiles it is possible to derive the detail level in the most recent representation of an object D_{object_i}, as well as the rate of change of the detail level, $\dot{D}_{\text{object}_i}$, as the explanation process continues. Both these quantities and the frequency of occurrence of an object within the profiles of the user model, F_{object_i}, are used by the Approximate Reasoner to compute the *desired detail level*, $\hat{D}_{\text{object}_i}$, of the fragments to be selected in the subsequent system model formulation process. Approximate reasoning [Pedrycz and Gomide, 1998] provides a consistent methodology of accounting for the uncertainty about the relationship between the user's expertise and the desired detail of the model fragments that is needed for domain model formulation.

The Model Formulator exploits the Approximate Reasoner's decision to select appropriate model fragments. It is this module that the current work is focused on. In CM, each of the model fragments typically represents a component or process under certain operational conditions and is associated with several consequences that will result if the conditions are satisfied. Given model fragments for various aspects of a physical system, the system model is formulated commonly by determining its boundary and deciding the best representation of parts within that boundary. For the present application, the boundary of the system model is set by an initial domain-dependent description of the system that the trainee is to learn about. The model formulation process is facilitated by reasoning with a Bayesian network. This process will be detailed in the following section. The outcome of this reasoning is a complete, adequate model containing fragments for all the components of interest, at a detail level appropriate for the user. The model is then parsed in order to extract information to be conveyed and the methods described in [Cawsey, 1993] are applied for structuring dialogue plans for this information to be communicated to the user via the presentation manager.

3 Model Formulation

After a general overview of the system's functioning, the details regarding the Model Formulator come into focus. Before engaging in the discussion of the main issues of the proposed model formulation technique, it is beneficial to briefly review the relevant theoretical aspects for the two research areas involved in the model formulation process, i.e. CM and Bayesian networks.

Compositional Modelling

CM offers a paradigm of formulating models for a given physical domain, composing them out of partial descriptions of certain physical aspects. These descriptions are generally known as *model fragments*. Each of these fragments represents a component or process under certain conditions and with several consequences which can be fulfilled provided that the conditions are satisfied.

Having represented a set of phenomena using model fragments, the methodologies developed within CM strive to complete two tasks: determining the model boundary and deciding the best representation of parts within that boundary. The former involves determining the boundary of the model with respect to both the physical extent of the system being modelled and the detail of representation of the various components of the system. Typically, a hierarchical manner of representation is followed, with all the components assembled by subcomponents. Traditionally, heuristics such as discovering the *minimal covering system* [Falkenheiner and Forbus, 1991] are used to identify the physical model boundary.

With the model boundary identified, a decision is made on how each part of the system should be represented as well as how to assemble a model from the partial descriptions for each of the parts. The main approaches are creating a parsimonious model directly, creating an over-complex model and then attempt to simplify it, or creating a simple model and subsequently adding detail to it according to the task for which the model is required.

The final issue to be addressed concerns the verification of the adequacy of the formulated models with respect to the task they are built for. The typical approach is to check the internal consistency of a model against a set of adequacy constraints [Levy *et al.*, 1997].

Compositional Modelling approaches are sound, but offer little provision for use in uncertain environments. In the context of explanation generation, however, selecting suitable model fragments with respect to the user's expertise inherently involves uncertainty in modelling assumptions. This is because of

- incompleteness in determining exactly what the user knows about the domain components of interest,
- approximation in specifying which level of detail can be deemed as sufficient when selecting fragments to satisfy a query,
- complexity in mapping all possible levels of expertise to all possible detail levels for the available model fragments.

A useful model formulation mechanism must be able to handle these issues. Bayesian network reasoning provides a general and efficient means of addressing uncertainty and is therefore used herein to assist the task of the Model Formulator. In addition the network structuring methodologies that accompany Bayesian networks can be an efficient basis upon which to translate modelling expertise into intuitive causal relationships, when selecting model fragments for different domain components.

Bayesian Networks

A Bayesian network is a directed acyclic graph in which the nodes represent the variables of interest and the links stand for the causal dependencies between variables (or nodes) [Pearl, 1988]. The nodes are labelled with conditional probabilities that provide estimates of the strength of the dependencies between the values of the variables. Therefore, a Bayesian network is a compact, localised representation of a probabilistic model, using a qualitative, graphical depiction of the causal relations between the entities involved, and a quantitative measure of the "strength" of these links. The key to its locality is that, given a graphical structure which represents the dependencies (and, implicitly, conditional independencies) among a set of variables, the joint probability distribution over the set can be completely described by specifying the appropriate set of marginal and conditional distributions over the variables.

Efficient reasoning mechanisms have been developed to handle inferences within a Bayesian network [Pearl, 1988], [Shennoy and Shafer, 1988]. Typically, these mechanisms involve propagation of received evidence for the value of a specific node throughout the network, using the Bayes rule [Pearl, 1988] for the estimation of posterior probabilities of the nodes taking specific values.

Several major issues are important to be resolved when applying Bayesian networks:

- *structuring the network*, based on qualitative information of the causal influences between the chosen representational primitives (i.e. the network's nodes);
- *eliciting probabilistic information* to annotate the network's links, i.e. deciding about the prior probabilities associated with the network's links; and
- *selecting a method to propagate evidence* throughout the network, and collecting results to utilise later on. These will be addressed below for the present application problem.

Other details relating to Bayesian networks are beyond the scope of this paper, but are discussed in depth in [Pearl, 1988].

The Proposed Method

The inputs to the Model Formulator are:

- A structural description of the domain system, specifying the components involved and the interconnections between them. This is prescribed before the start of an explanatory session and remains unchanged during the explanation process.

- A library of model fragments, represented at different detail levels. Each model fragment represents a certain aspect of the structural and behavioural information of a given component of the domain system. The model fragments for the different representations of a particular component are organised into an *assumption class* for the component. Fragments within an assumption class have different and often mutually contradictory conditions, and thus only one fragment from the assumption class of a given component can be used to represent the component in a system model.
- A desired level of detail for a subset of the model fragments required to construct the system model, provided by the Approximate Reasoner.
- A set of objects (components or processes) of interest, obtained from the queries of the user.

The domain system description is used to structure a Bayesian network relating model fragments from one assumption class to fragments of other classes. Each network node stands for the selection or rejection of a specific model fragment. As such, a node may take a value from the set $\{yes, no\}$ (with *yes* indicating the selection of the fragment and *no* the rejection). The links of the Bayesian network represent the relationships between model fragments as determined by the global system description: if component A has its output connected to the input of component B, then links are established from the model fragments of A towards those of B. Although care needs to be taken over the "connections" between fragments of different precisions, the model fragments used are of qualitative symbolic nature and such connections can be reasoned on the grounds of qualitative values of the variables involved in the fragments. The present work assumes the worst case, allowing all model fragments of component A to be connected to all fragments of component B, if A is physically connected to B. Of course, any given constraints from theoretical or empirical knowledge sources that prohibit such a connection can be taken into account in setting up the network structure by assigning a prior probability of zero to the forbidden network links.

Given the structure of a Bayesian network, each network node is annotated with estimates for the conditional probabilities of the node acquiring a *yes* or *no* value when its "parent" nodes are assigned their values. In general, deriving such prior probabilities is a task of great difficulty [Pearl, 1988]. There are applications where historical data is available, thereby enabling the required estimation, sometimes by simply calculating the frequencies of value appearance. In less fortunate cases, like the present application, some sort of rules or heuristics have to be derived from the problem description in order to decide the prior probabilities.

In this work the heuristics employed depend on whether or not a node is a root node, i.e. one that has no parent nodes. The first heuristic indicates whether a model fragment is to be selected, by weighing the following two important factors: the strength of causal influence that the fragment receives from its selected parents and the compliance of the fragment with respect to the detail level of these parent nodes. This heuristic can be stated as follows.

Definition 1 *For each non-root node MF_{j_i}, $j = 1, \ldots, L$, representing the selection or rejection of model fragment j of component C_i, $i = 1, \ldots, M$, with parent nodes $U_{jk}, k = 1, \ldots, N$, the prior probability of selecting the corresponding fragment when some of the parent nodes are taking a value yes (with the remaining parent nodes taking a no value) is determined by:*

$$\begin{aligned} P(MF_{j_i} = yes | \textstyle\bigcap_p U_{jp} = yes) &= \\ & P(MF_{j_i} = yes | \textstyle\bigcap_q U_{jq} = yes, D_{U_{jq}} = D_{MF_{ji}}) \\ & \cdot P(MF_{j_i} = yes | \textstyle\bigcap_r U_{jr} = yes, D_{U_{jr}} \neq D_{MF_{ji}}) \end{aligned}$$

with

$$\begin{aligned} P(MF_{j_i} = yes | \textstyle\bigcap_q U_{jq} = yes, D_{U_{jq}} = D_{MF_{ji}}) & \\ = \quad \textstyle\prod_q V_{MF_{ji} \leftarrow U_{jq}} & \\ P(MF_{j_i} = yes | \textstyle\bigcap_r U_{jr} = yes, D_{U_{jr}} \neq D_{MF_{ji}}) & \\ = \quad \frac{\prod_r V_{MF_{ji} \leftarrow U_{jr}}}{\prod_r (D_{MF_{ji}} - D_{U_{jr}})^2} & \end{aligned}$$

In this definition, p ranges among the parents that have taken a *yes* value, and $V_{MF_{ji} \leftarrow U_{jl}}, l \in \{q, r\}$ denotes an estimate of the amount of causal influence that fragment MF_{j_i} receives from parent U_{jl} when U_{jl} takes a value *yes*. Let $N_{influence}$ be the number of those variables which are defined in the consequences of U_{jl} and which are influencing variables of MF_{j_i}, and N_{total} be the total number of variables defined in U_{jl}, then $V_{MF_{ji} \leftarrow U_{jl}}$ is calculated as the ratio $N_{influence}/N_{total}$.

As an example of the heuristic, consider the simple network of figure 2 which connects four model fragments of different detail levels (for three components A, B and C). Model fragment $A1$ has one variable out of a total two variables that can influence variables in its child node, fragment $B1$, that is $V_{B_1 \leftarrow A_1} = 0.5$. The "error" in detail between fragments B_1 and A_1 is $D_{B_1} - D_{A_1} = 4 - 2 = 2$. Similarly, $V_{B1 \leftarrow A_2} = 0.33$, $V_{C_1 \leftarrow B_1} = 0.50$, $D_{B_1} - D_{A_2} = 1$ and $D_{C_1} - D_{B_1} = 0$. The application of the above heuristic gives, thus, the following non-normalised results:

$$\begin{aligned} P(B1 = yes | A1 = yes, A2 = yes) &= 0.0413 \\ P(B1 = yes | A1 = yes) &= 0.125 \\ P(B1 = yes | A2 = yes) &= 0.33 \\ P(C1 = yes | B1 = yes) &= 0.50 \end{aligned}$$

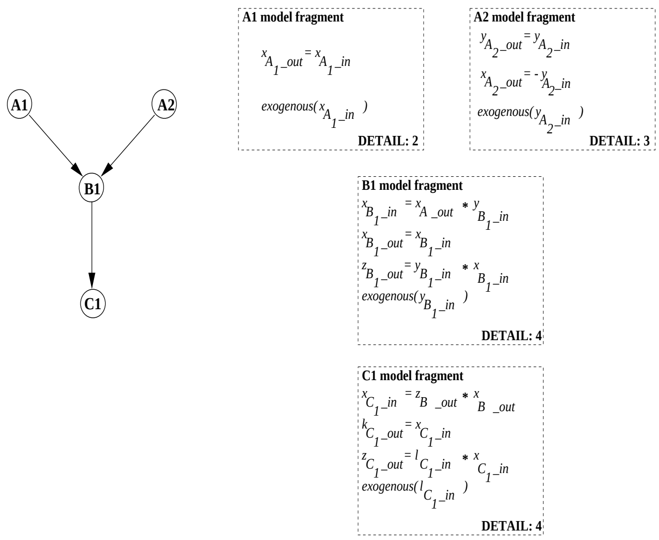


Figure 2: *A simple Bayesian network* and the corresponding model fragments

A much simpler heuristic is used to assign prior probabilities to root nodes, assuming all fragments of a component are initially equally likely to be selected.

Definition 2 *For all root nodes RMF_{r_i}, $r = 1, \ldots, R$ of a component C_i, $i = 1, \ldots, S$, the prior probability of them being selected is $P(RMF_{r_i} = yes) = \frac{1}{R}$.*

Following this, for the root nodes $A1$ and $A2$ of figure 2, their probability of being selected is:

$$P(A1 = yes) = P(A2 = yes) = 1/2$$

Given a subset of the prior probability values for the network's root nodes and for each non-root node with respect to its selected parents, the probabilities for the remaining node values combinations can be calculated. This is accomplished using the principle of *maximum entropy*, subject to the constraint that the sum of all probability values to a node must be equal to 1 and assuming a uniform distribution of those prior probabilities for which we have no previous estimation (see [Pearl, 1988] for details of this principle).

Given the structure and prior probabilities, reasoning is performed based on the evidence of which fragments of what components are of current interest to the user and on which model fragments are indicated by the Approximate Reasoner. The reasoning process helps with the decision of which fragments to select (see figure 1). The evidence is translated to selection of the indicated fragments with probability 1. This is propagated throughout the network to determine the posterior probabilities of the nodes for which there exists no evidence, using a standard Bayesian network inference method as detailed in [Shennoy and Shafer, 1988]. When the reasoning process terminates, the posterior probabilities of every model fragment being selected or rejected are returned.

Model acceptance under adequacy constraints

The Bayesian reasoning process determines the most probable representation of each individual component with respect to the user's expertise. Although the reasoning is mathematically sound, ensuring that only one fragment is chosen for each component, the system model constructed by simply combining the selected model fragments through their terminal connections is not entirely guaranteed to be most adequate in supporting the user's information needs. Additional constraints may be needed to modify the decisions made by Bayesian inference to eliminate inclusion of fragments that would have led to an inadequate model. The criteria listed in table 1 are used to serve as the adequacy constraints, the intuitive knowledge of human modelers.

These constraints are applied whenever a model fragment is considered for selection (or rejection) according to the reasoning of the network. The partially formulated model is checked for its adequacy, and if not all constraints are satisfied, the model fragment is discarded. In this case, the next most probable model fragment (of the same component) is considered as a candidate for selection. A fragment cannot be retracted once it has passed the test of the adequacy constraints. This is to allow for consistent application of the adequacy constraints avoiding potential loops.

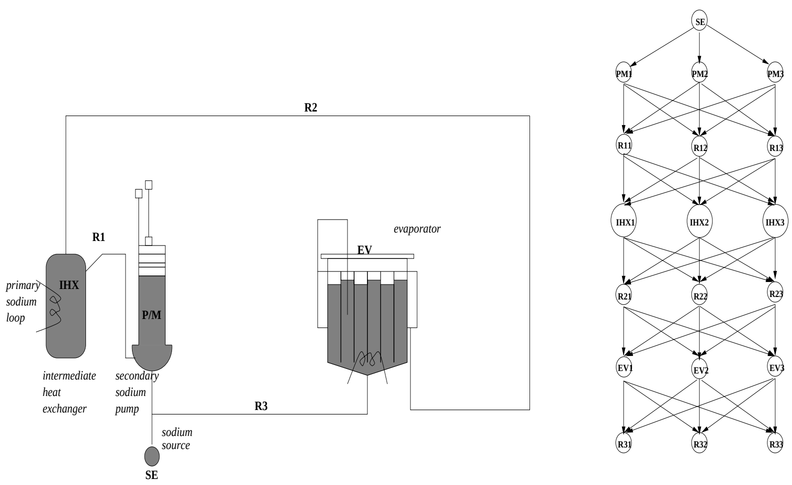
4 Example of formulating adequate models

To demonstrate that the proposed model formulation technique works, the approach has been implemented and applied to the domain system depicted in figure 3. The example system is a simplified version of the secondary liquid sodium cooling system within a fast breeder nuclear power plant. It contains seven components: IHX (intermediate heat-exchanger), P/M (pump-motor system), EV (evaporator), SE (source of liquid sodium), and R1, R2 and R3 (hydraulic resistances of the connecting pipes). For simplicity, only hydraulic phenomena are considered for each of these components. The model fragments library contains three model fragments with detail levels of 4, 6 and 10 variables for each component, except the liquid sodium source which is modeled with one model fragment with a detail level of 4 variables. The fragments for each component are grouped using the component's assumption class. Suppose that each component within the system is influenced only through its two input terminal variables, the input liquid flow and the input pressure. Consequently, each fragment can influence other fragments through its output terminal variables: the output pressure and the output flow. This is the worst scenario with respect to the possible relations between variables defined in system: knowing practically nothing about the causal relationships between the variables.

Given the domain system description, the Bayesian network can be structured as also shown in figure 3. Using the heuristics defined in section 3, the prior prob-

No.	*Primitives involved*	*Description*	*Rationale*
1	Model variables System components	An adequate model should include every component of interest.	The behaviour of a component cannot be explained if the component is not represented in the model.
2	Exogenous variables	In an adequate model none of its exogenous variables should be influenced by any other model variable.	Exogenous variables are influenced only by the surrounding environment of the modelled system and are independent of other variables in the model.
3	Influences on dependent variables must be complete	An adequate model should contain a *complete* set of model fragments that influence each of the model's dependent variables, according to the overall domain system description.	This ensures that a model is complete: for each component the model should include model fragments for all components that may affect it, at the detail level of interest, according to the domain system description.
4	Influences on dependent variables must not be redundant	An adequate model should contain no fragments that relate to each other through class inheritance (e.g. a condenser and a heat exchanger fragment for one component).	This is imposed to ensure that when composing model fragments together coherence is maintained by avoiding to mix different description levels of an entity.
5	Influences on dependent variables must be valid	An adequate model should include model fragments with valid conditions.	This ensures that all selected model fragments have valid conditions (structural and behavioral) with respect to the entities appearing in the model at the respective detail level.

Table 1: *Adequacy constraints:* description and rationale



SE

T00	
y	*1.0*
n	*0.0*

Prior probability of selecting the unique model fragment for component "SE"

PM1 (SE)

T11	*y*	*n*
y	*0.88*	*0.04*
n	*0.04*	*0.04*

PM2 (SE)

T12	*y*	*n*
y	*0.52*	*0.16*
n	*0.16*	*0.16*

PM3 (SE)

T13	*y*	*n*
y	*0.1*	*0.3*
n	*0.3*	*0.3*

Conditional probability tables for the model fragments "PM1" "PM2" "PM3" with respect to the selection or not of the model fragment for "SE"

R11 (PM1 & PM2 & PM3)

T21	*yyy*	*yyn*	*yny*	*ynn*	*nyy*	*nyn*	*nny*	*nnn*
y	*0.0004*	*0.0605*	*0.0041*	*0.7323*	*0.0007*	*0.1209*	*0.0082*	*0.0081*
n	*0.0081*	*0.0081*	*0.0081*	*0.0081*	*0.0081*	*0.0081*	*0.0081*	*0.0081*

R12 (PM1 & PM2 & PM3)

T22	*yyy*	*yyn*	*yny*	*ynn*	*nyy*	*nyn*	*nny*	*nnn*
y	*0.0008*	*0.0711*	*0.0028*	*0.2152*	*0.0071*	*0.5681*	*0.0215*	*0.0126*
n	*0.0126*	*0.0126*	*0.0126*	*0.0126*	*0.0126*	*0.0126*	*0.0126*	*0.0126*

R13 (PM1 & PM2 & PM3)

T23	*yyy*	*yyn*	*yny*	*ynn*	*nyy*	*nyn*	*nny*	*nnn*
y	*0.0002*	*0.0009*	*0.0081*	*0.0404*	*0.0119*	*0.0598*	*0.5808*	*0.0331*
n	*0.0331*	*0.0331*	*0.0331*	*0.0331*	*0.0331*	*0.0331*	*0.0331*	*0.0331*

Conditional probability tables for the model fragments "R11" "R12" "R13" with respect to the selection or rejection of the combination of model fragments "PM1" "PM2" PM3"

Figure 3: *Liquid sodium cooling loop*, its corresponding Bayesian network and some prior probabilities

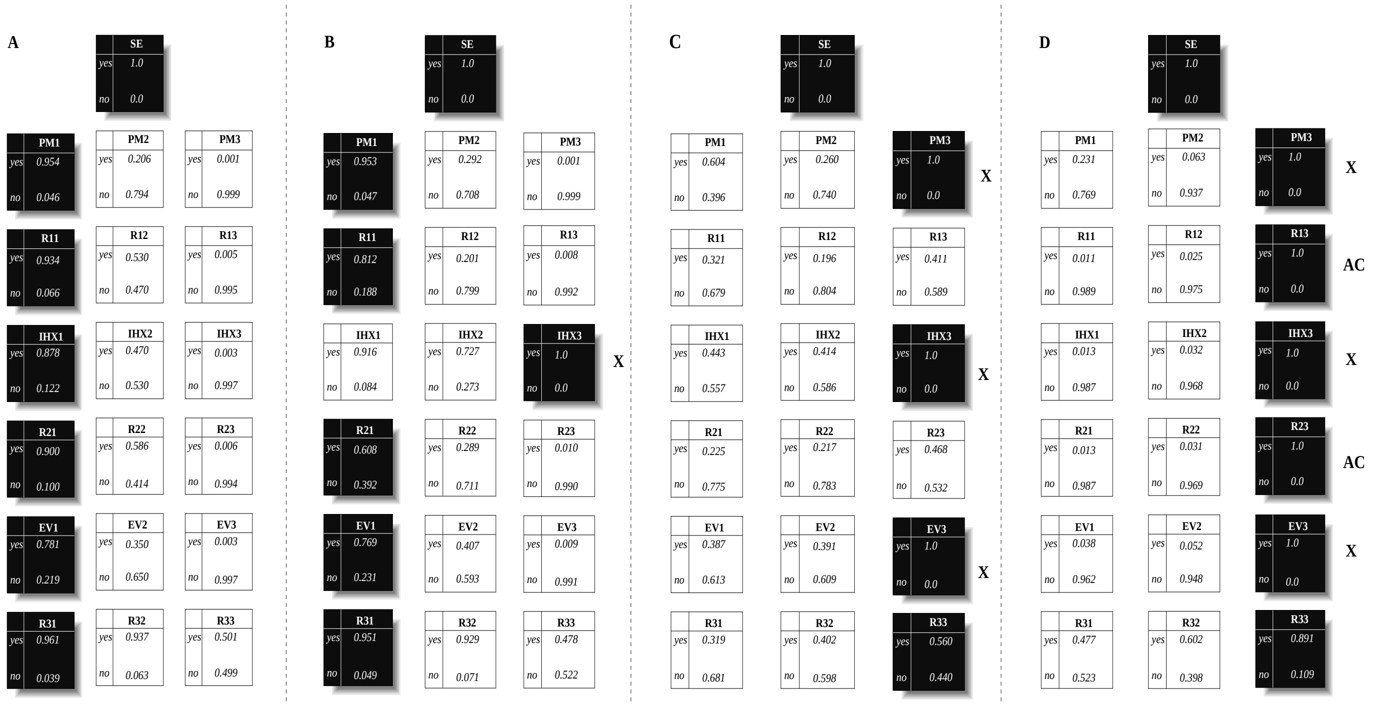


Figure 4: *The posterior probability distribution for the network in figure3. A: without* considering evidence; *B:* with IHX3 selected by evidence; *C:* with PM3 and EV3 selected by evidence; *D:* with the selection of R13 and R23 biased through application of adequacy constraints.

abilities for each of the nodes with respect to the links from its parents, as well as the prior probabilities for the root node (SE) are determined. Part of the setting of the normalised prior probabilities is presented in figure 3. The prior probability value tables attached to the remaining nodes are the same as those listed in tables T21-T23 in figure 3 and hence are omitted. Once the network is structured and the prior probabilities are assigned and normalised, the inference procedure reported in [Saffiotti and Umkehrer, 1991] is ready to be applied to calculate the posterior probabilities for possible values of the network nodes.

Without considering evidence from the user or the Approximate Reasoner, the posterior probability distribution of the Bayesian network is computed as that given in sample *A* of figure 4. The shaded mini tables in this figure correspond to the elements that are indicated for selection, as a direct result of Bayesian reasoning. Given no evidence, the simplest model fragments are selected, meeting the training purposes well.

Suppose now that the user has indicated interest for the intermediate heat exchanger component. This is translated to a piece of evidence suggesting that a model fragment of this component should be selected. Assuming that the Approximate Reasoner has calculated a detail level of 10 variables for this component, the most complex model fragment is selected. This evidence is propagated throughout the network, resulting in the probability configuration for other system components as shown in sample *B* of figure 4, with the evidence denoted by *X*. As indicated the simplest fragments are still favoured over the complex ones, although the probabilities of selecting complex model fragments have increased compared to those in sample *A*.

If the indications for the preferable model fragments continue to favour the most complex ones, as with sample *C* of figure 4 where the fragments for PM and EV are also selected by given evidence, this can lead to a change of the fragment selection. However, as illustrated by this example, although the most complex fragments are in favour compared to the simpler ones, some of them obtain a posterior selection probability that is less than 0.5. That is, the corresponding fragments should not be selected in a strict mathematical sense. Therefore, the inference over the Bayesian network can indeed help identify which model fragments to select, but it cannot guarantee that *all* components in the system under consideration will be represented. The prior network probabilities are responsible for this: they are biased towards simpler model fragments, causing the probability of the complex fragments to be less than 0.5, as simple fragments are not supported by the incoming evidence. This gives rise to the need of imposing the adequacy constraints.

In fact, the adequacy of the model under formulation is checked every time a model fragment is about to be selected or rejected for a component. For this particular example, when the network attempts to overrule all existing fragments for a component, adequacy constraint no. 3 is violated. This triggers a mechanism that overrides the network's decision, by asserting as evidence the selection of the most probable fragment among the otherwise overruled ones. The reasoning process of the network restarts thereafter. The amended result for the cases of components R1 and R2 is shown in sample D of figure 4, with the selections imposed by the use of adequacy constraints denoted as AC. The resulting system model is well-suited for generating explanations that fit the expertise level of the present trainee.

5 Conclusions

This paper has proposed a technique for model formulation to support the task of explanation generation. The technique exploits the reasoning of a Bayesian network, which is structured based on the structural description of the domain system, to facilitate the selection of appropriate model fragments. Initially, fragments for some components are selected, based on the user's interests and expertise level. Bayesian reasoning provides suggestions for model fragments to use for the remaining part of the system. The model under formulation is then checked for its adequacy, using a set of adequacy constraints. The final result is an adequate model of the domain system under consideration, which can be subsequently analysed to extract contents for the explanations to be communicated to the user.

The proposed approach has been implemented and experimental results obtained so far have been very promising. The methodology described employs the simplest regimes for addressing the issues of structuring the Bayesian network and those of defining the prior probabilities for the task of model fragment selection. Although it functions satisfactorily for simple cases, it needs to be re-engineered in order to become more general and less ad-hoc with respect to the use of the adequacy constraints. Work is also ongoing in an attempt to automatically construct models using the present approach for more complex systems.

References

[Cawsey, 1993] A. Cawsey. *Explanation and Interaction.* MIT Press, 1993.

[Falkenheiner and Forbus, 1991] B. Falkenheiner and K. Forbus. Compositional Modeling: finding the right model for the job. *Artificial Intelligence*, 51:95–143, 1991.

[Gruber and Gautier, 1993] T. R. Gruber and P. Gautier. Machine-generated explanations of engineering models: A compositional modelling approach. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pages 1502–1508, 1993.

[Levy *et al.*, 1997] A. Y. Levy, Y. Iwasaki, and R. Fikes. Automated model selection for simulation based on relevance reasoning. *Artificial Intelligence*, 96:351–394, 1997.

[Nayak, 1994] P. P. Nayak. Causal approximations. *Artificial Intelligence*, 70:277–334, 1994.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan-Kaufmann, 1988.

[Pedrycz and Gomide, 1998] W. Pedrycz and F. Gomide. *An Introduction to Fuzzy Sets: Analysis and Design.* MIT Press, 1998.

[Saffiotti and Umkehrer, 1991] A. Saffiotti and E. Umkehrer. Pulcinella: A general tool for propagating uncertainty in valuation networks. In *Proceedings of the 7th Conf. on Uncertainty in AI*, pages 323–331, 1991.

[Shennoy and Shafer, 1988] P.P. Shennoy and G. Shafer. An axiomatic framework for bayesian and belief-function propagation. In *Proceedings of AAAI Workshop on Uncertainty in AI*, pages 307–314, 1988.