

Using Bayesian Networks as an Inference Engine in KAMET*

Osvaldo Cairó and Rafael Peñaloza
Instituto Tecnológico Autónomo de México (ITAM)
Department of Computer Science
Río Hondo 1, Mexico City, Mexico.
cairo@itam.mx, rpenalozan@yahoo.com

Abstract

During the past decades, many methods have been developed for the creation of Knowledge-Based Systems (KBS). For these methods, probabilistic networks have shown to be an important tool to work with probability-measured uncertainty. However, quality of probabilistic networks depends on a correct knowledge acquisition and modelation.

KAMET¹ is a model-based methodology designed to manage knowledge acquisition from multiple knowledge sources [1] that leads to a graphical model that represents causal relations. Up to now, all inference methods developed for these models are rule-based, and therefore eliminate most of the probabilistic information.

We present a way to combine the benefits of Bayesian networks and KAMET, and reduce their problems. To achieve this, we show a transformation that generates directed acyclic graphs, the basic structure of Bayesian networks [2], and conditional probability tables, from KAMET models. Thus, inference methods for probabilistic networks may be used in KAMET models.

1. Introduction

Probabilistic networks have shown to be an important tool for Knowledge-Based Systems (KBS) that include, within their knowledge, probabilistic-measured uncertainty, and relational information about the domain. In particular, if the network model may be represented by a directed acyclic graph (DAG), and includes discrete variables, the use of Bayesian networks is specially adequate. Several methods have been developed to carry out efficiently exact inferences in some particular cases of probabilistic networks (which

include Bayesian networks); even more, if some error is tolerated, there are in the literature many methods of approximate inference that reduce considerably the machine time needed to solve the problem [7]. In the last years, probabilistic networks have shown to have very important applications, far beyond the theoretic ground (see, for example, [5]).

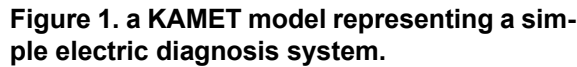
The quality of the probabilistic models depends strongly on the correctness of the causality and independence representations. Unfortunately, there has been little progress in the creation of methodologies that create correct probabilistic models. The problem of creating such a representation is equivalent to acquire and represent expert knowledge in an appropriate way. KAMET (Knowledge Acquisition Methodology) is a methodology based on models designed to manage knowledge acquisition, mainly for diagnosis, from multiple knowledge sources. As a result of this methodology, one obtains a probabilistic model that represents causal relations between the possible symptoms and problems, with an inaccuracy associated to each relation.

An issue that has followed KAMET right from the beginning is to find the way to formalize those models, so that they allow to recognize appropriate causes to the problems given at the time. In this article, two methods for transforming KAMET models into Bayesian networks are proposed. The reason for presenting two transformations instead of just one is that each has benefits, one in the model representation and other in the size of the Bayesian network obtained. These transformations look different at first sight, however, it can be shown that the marginal distributions over the common nodes in both transformations will be identical. Thus, exact inference methods will yield to the same results. To make more clear these transformations and their definitions, we will use a simple KAMET model that represents an electric diagnosis system, shown in figure 1.

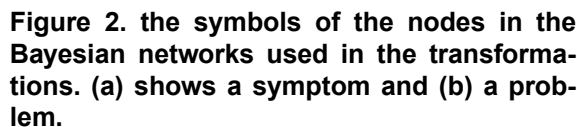
This model will be explained in section 2, after a brief description of the KAMET CML.

* This project has been founded by CONACYT, as project number 33038-A, and Asociación Mexicana de Cultura, A.C.

¹ KAMET is a project that is being carried out in collaboration with the SWI Group at Amsterdam University and Universidad Politécnica de Madrid.



In order to keep the additional benefit obtained by KAMET, making symbolic distinctions between the elements in which the diagnosis is based, nodes representing symptoms will be diamond shaped, and those related with problems will have the shape of a square, as shown in figure 2, even when the Bayesian net-



The KAMET CML has three levels of abstraction. The first one corresponds to structural constructors and structural components. The structural constructors are used primarily to highlight the problem itself. These are: *problem*, which expresses an alteration, disorder or abnormality, *classification*, alterations disorders or abnormalities that can be considered as a classification problem and, thus, may be

works will make no difference between them. We will give these nodes the same name they had in the origi-

represented within a table, and *subdivision* of alterations, disorders or abnormalities (figure 3).

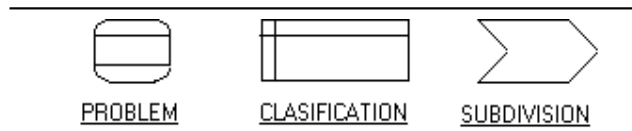


Figure 3. structural constructors.

The structural components are used to establish the characteristics and possible solutions of the problem (figure 4). They have the following meaning:

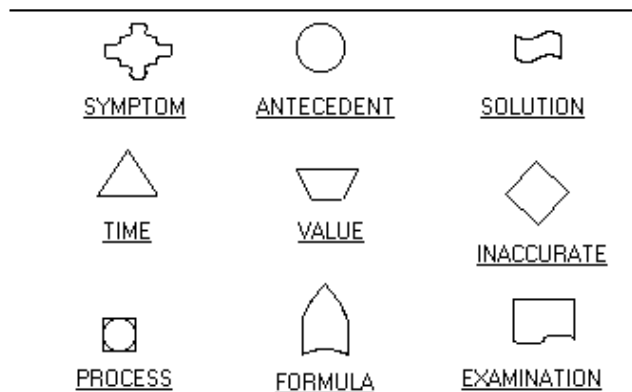


Figure 4. structural components.

- Symptom is a manifestation or sign related to an alteration, disorder or abnormality.
- The antecedent expresses the previous circumstances that can be used to judge something that might happen.
- A solution expresses the possible solution to a disorder, alteration or abnormality. It is always related to structural constructors.
- Time expresses the duration of structural components and structural constructors.
- A value expresses the characteristics of symptoms, antecedents or groups.
- Inaccurate expresses the uncertainty or lack of precision of an intermediate or terminal node.
- The process expresses the sequence of actions or series of operations required to obtain a result.

- A formula expresses the calculations that must be compelled in order to determine the alteration, disorder or abnormality.
- Examination expresses a recommendation or necessity of making studies to determine an alteration, disorder or abnormality.

The second level of abstraction corresponds to *nodes* and *composition rules*. Models are represented by digraphs, with nodes as their vertices and composition rules as arcs. Structural constructors and structural components form the nodes, which are divided in three kinds: initial, intermediate and terminal. Composition rules permit the adequate combination of structural constructors and components (figure 5):

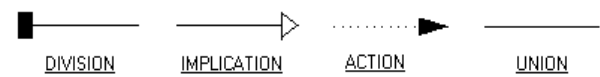


Figure 5. composition rules.

- Division expresses that an alteration, disorder or abnormality is subdivided in ■
- The implication expresses connection from an origin toward a complication.
- Action expresses that something must be completed (a formula or an examination).
- Union expresses connection between subdivisions.

The third level of abstraction corresponds to the global model, which should represent the knowledge acquired from multiple knowledge sources. It must have at least one initial node and one terminal node. In order to facilitate the construction and interpretation of the models, structural constructors and components may be *named* using a numerical or linguistic label.

Another element related to the models is the indicator, represented by a square (figure 6), located in the upper right-hand corner of the group or the structural component value being referred to. An indicator is named in three different ways: the notations n , $n+$, and n, m (with $m > n$) are used to express the exact, minimum, and minimum and maximum number of elements that must be present, respectively.

A *group* is a set of linked elements that have times and/or values in common, or are related among them through an indicator; a group may be an element of another group (figure 7). A *chain* is the link of two or more groups, symptoms and/or antecedents. In the graphical representation, the order of the link is irrelevant.

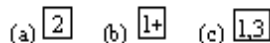


Figure 6. indicators. (a) indicator n ; (b) indicator $n+$; (c) indicator n, m .

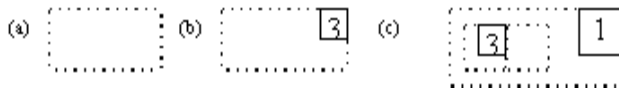


Figure 7. groups. (a) group; (b) group with indicator; (c) recursive group.

An initial node represents a symptom, antecedent, group or chain, and does not have input flows. An intermediate node represents a structural constructor or any of the structural components: study, process, formula, or inaccuracy. Finally, a terminal node represents a structural constructor; it has at least one input flow, and their output flows are used only to determine possible solutions. The initial and intermediate nodes can be grouped together, without losing their properties or functions, into molecular nodes, by means of conjunctions or disjunctions (figure 8). These nodes will, in

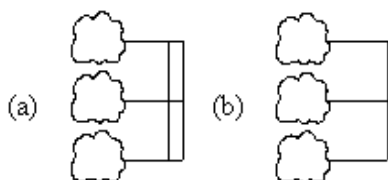


Figure 8. molecular nodes. (a) conjunction and (b) disjunction.

turn, act as nodes in their own right.

For example, the model in figure 1 expresses that the problem P1 can occur due to two different situations. In the first one, the model expresses that if symptoms 1 and 2 are known to be true, then we can deduce that problem P1 is true with probability 0.7. In the second one, the model expresses that if symptoms 1 and 5 are observed then we can conclude that the problem is P1 with probability 0.6. On the other hand, we can deduce that the problem P3 is true with probability 0.4 if symptoms 1 and 4 are known to be true. Finally, we can reach a conclusion that the problem is P2 with probability 0.9 if problems P1 and P3 and the symptom 3 are observed.

The prior probabilities of the nodes have no symbolic representation in KAMET, but they may still be known; in that case, they are stored in a table. For a more complete description of KAMET, see [1].

3. Transformation with AND nodes

An essential part of KAMET models are their molecular nodes, representing the conjunction and the disjunction. The first transformation proposed uses the **AND** connector², creating a new type of node which does not represent symptoms nor problems; this will be called an **AND** node, and its representation will be by a triangle, as shown on figure 9. These nodes will have an upper case letter as a name. Note that we won't create a new node for disjunctions because, in general, this would lead to the loss of information on the conditional probabilities of the new **OR** node given each of its parents.



Figure 9. an AND node called T.

The Bayesian network is constructed generating a node for each component in the KAMET model. An **AND** node is then generated for each conjunction, and a directed edge from each component in it to the **AND** node is created. The conditional probability table is the trivial one, given by

$$P(\text{AND} | N_1 = n_1, \dots, N_k = n_k) = \prod_{i=1}^k n_i$$

where $N_j, j = 1, \dots, k$ are the parents of the **AND** node; thus, the **AND** node has a value of 1 if and only if each of its parents has a value of 1³. To finish the transformation, a directed edge is created from the **AND** node to the node pointed by the implication.

If there exist **OR** connectors, directed edges are created from those components joined by a connector of this kind to the node pointed by the corresponding implication.

After applying this transformation to the model in figure 1, the Bayesian network with **AND** nodes obtained is that shown in figure 10.

We now need to generate the conditional probability tables. The first case is when a node N has only one parent.

- 2 We will use interchangeably the phrases “**AND** connector” with “conjunction molecular node”; and “**OR** connector” with “disjunction molecular node”.
- 3 We use the convention that 1 means that the node has a value of TRUE (it is observed), and 0 means that it's FALSE.

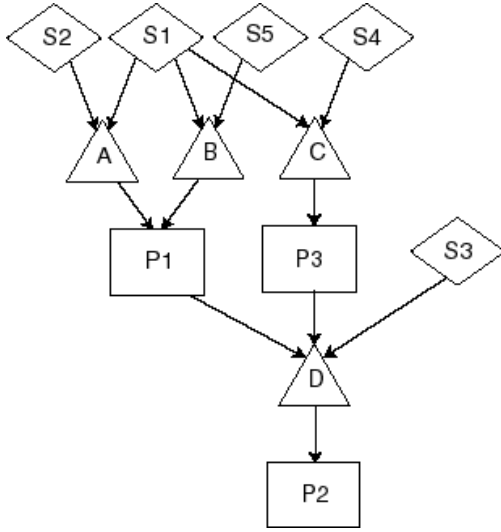


Figure 10. transformation of the KAMET model shown before using AND nodes.

We then know the conditional probability of the node given that we observed that its parent appeared $P(N|pa(N))$. However, the model has no information over the conditional probability given that its parent is observed not to appear; in this case, we may fill the table with the prior unconditional probability of N , if we don't have more information about the node.

The other case is when N has more than one parent. By construction, the parents of N must form a disjunctive molecular node. The probability table will have the value of the uncertainty associated to one of the parents if and only if that parent has a value of true and none of the others. For the other cases (none or more than one of the parents has a value of true), the prior probability of node N may be used. In some cases it's possible to make other kinds of approximations; for example, if two of the parents are known to be mutually exclusive, then putting a value of 0 to the probability of the node given both parents is useful in the propagation stage. On the other side, it is common to suppose that if more symptoms are known to be present, the probability of a problem should be higher; thus, one may insert values higher to those given to each parent individually.

For the example we are working with, let the problems have prior probabilities given by

$$\begin{aligned} P(P1) &= 0.25 \\ P(P2) &= 0.1 \\ P(P3) &= 0.2 \end{aligned}$$

and suppose that we have additional information that suggests that if we find symptoms $S1$, $S2$, and $S5$, the probability of occurrence of problem $P1$ is 0.1 higher than the

one given when only symptoms $S1$ and $S2$ are present. We will also suppose that all the symptoms have a prior probability of 0.5. The conditional probability tables associated to the nodes in figure 10 are shown below. We omit the tables of the **AND** nodes.⁴

	P1	¬P1
A,B	0.8	0.2
A,¬B	0.7	0.3
¬A,B	0.6	0.4
¬A,¬B	0.25	0.75

	P3	¬P3
C	0.4	0.6
¬C	0.2	0.8

	P2	¬P2
D	0.9	0.1
¬D	0.1	0.9

This transformation has the advantage that maintains the representation given by the original KAMET model, with a structure of the same kind; therefore, the relationship between both models is clear. However, it tends to add many nodes, and the resulting graph will be very large and will form many cliques during the moralization process; thus, the propagation and inference methods may become very slow.

4. Direct transformation

The direct transformation eliminates the problem of generating models that are too big, because it doesn't create new nodes, using those that appear in the original model. This transformation can be performed directly, in stead of making many different steps as in the other transformation.

The first step is to generate the nodes that will be in the graph, one for each symptom or problem, and assign them a name (as in the past section). Each node will have as parents all the nodes that incide over it in the original model, without paying attention to the connectors that unites them. In the example we are following, the result of the transformation is given by figure 11.

After that, the conditional probability tables of each node given its parents are calculated using the information contained in the connectors of the original model. If there are components united by an **AND** connector, then the corresponding probability will be given only if all of them (and only them) occur.

⁴ The symbol $\neg A$ means that the node A didn't appear (has a value of false).

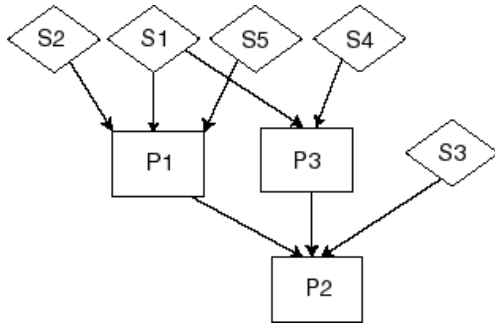


Figure 11. the direct transformation of the KAMET model of electric diagnosis.

Using the same information as in the other transformation, we obtain the following probability tables.

	P1	¬P1
S1,S2,S5	0.8	0.2
S1,S2, ¬S5	0.7	0.3
S1, ¬S2,S5	0.6	0.4
otherwise	0.25	0.75

	P3	¬P3
S1,S4	0.4	0.6
o.w.	0.2	0.8

	P2	¬P2
P1,P3,S3	0.9	0.1
otherwise	0.1	0.9

The prior probability of occurrence of the problems is used in many cells of the table because we don't know what happens when we do not observe all the elements united by an **AND** connector. Here one may guess that the inference in both models will produce different outcomes; for example, if the conditional probability of the problem P1 given the symptom combination $\neg S1, S2, S3$ was known (and was different to 0.25), the conditional probability table will be different. However, if that probability was known, it will be represented in the original model, and the transformation using **AND** nodes will include it in its representation.

A problem of the direct transformation is that, although it is pretty natural, produces a loss of the graphical information contained in the KAMET models. When **AND** nodes are used, it is easy to find out the relationship between the parents of a given node, so that one is able to know which symptoms should appear together in order to be able to say something about the problem; in the direct transformation, when we omit the use of this kind of nodes, the graphical model obtained shows a relation of the same kind between all the parents of a node, even when in the original model that is not true. This problem may be partially solved with the observation of the conditional probability tables. How-

ever, a clear graphical representation is fundamental in big problems, where an inspection of all the probability tables is impossible.

So, each transformation has its advantages and disadvantages compared to the other; the choice of one or the other depends on the kind of representation wanted with it, and on the size of the original model.

5. Equivalence

We won't show a complete proof of the equivalence of the Bayesian networks obtained by both transformations. Instead, we will give a motivation for this affirmation, in order to make it more clear.

First, by definition, an **AND** node will have a value of "true" if and only if all of its parents have the same value. Also, if an **AND** node performs as a separator in the Bayesian network of the first transformation, then in the directly formed Bayesian network, the set of its parents will perform as a separator too; this is also true backwards. Thus, if any of the transformation-based Bayesian networks satisfies a Markov property, then the other one will also satisfy the same; so, both networks may use the inference methods.

Also, the prior marginal probability of every node, except of course the **AND** nodes, is the same in both networks, so if one adds information of the appearance of a symptom or a problem, the marginal probabilities after propagation of the information will still be the same in both models. If the information we receive corresponds to the observation of an **AND** node, then it is equivalent to the observation of all of its parents, thus we may include this information in the model formed by the direct transformation.

6. Generalization

Up to now, we have supposed that all the nodes in the original KAMET model represent binary variables. Nevertheless the same transformations may be generalized to use variables which may take a finite number of different values, shown by the structural constructor value, even if it comes with an indicator. The generalization may not always be used, though, as we'll see next.

This generalization is made the natural way: the graphical representation of the transformations will be the same, and the difference will appear only in the tables of conditional probability, where all the possible combination of values of each of the parent nodes will appear. Within this table, one must be able to represent the probability of the problem given that an antecedent is not present; for that, we must add another value (typically '0') to represent the absence of all other values. Thus, if a node has k values, then the probability table of a node given this lone parent will

have 2^{k+1} entries for each possible value of the son; thus, models in which nodes have many values may become intractable for the size of its tables. However, depending on the indicator associated to the values, many of the combinations may not be present, reducing the size of the conditional probability table.

One important feature of this generalization is that it is independent of the arguments given in the past section for the equivalence. This means that even if the network come from the generalized transformation, these will still be equivalent; so there is no problem in using any of the transformations.

Notice that a node with multiple values may be seen as a conjunctive molecular node in which each of its atomic nodes represents a subset of the values it may take. With that in mind, the process of filling in the conditional probability tables when multiple-valued nodes are involved has no difference with filling the probability tables in sons of conjunctive molecular nodes. This reasoning may be useful in order to find a way to reduce the size of conditional probability tables in nodes with a large number of values.

7. Conclusions

We have shown two ways to obtain a Bayesian network from a KAMET model. With these transformations, we have solved one of KAMET's problems, the lack of inference methods, taking advantage of the well studied inference methods developed for Bayesian networks. By means of these transformations, one is able not only to use inference methods in KAMET models, but also to apply any Bayesian network-based methodology that may be developed. We didn't present a proof for the equivalence, up to inference methods, of these transformations, but it is easy to show it, because the marginal densities are equal in both models.

Current research in KAMET focuses on developing a new generalization for nodes that have many values, or inference methods that perform efficiently in Bayesian networks that come from this generalization, such that inference problems may be solved in a reasonable time. Research also focuses on implementing these transformations in order to be able to obtain Bayesian networks from graphical representations of the knowledge acquired by KAMET, and to implement some inference methods (exact and approximate) as Problem Solving Methods.

References

- [1] Cairó, O. (1998). *The KAMET Methodology: Content, Usage and Knowledge Modeling*. In Gaines, B. & Mussen, M. (Eds.). *Proceedings of the 11th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'98)*. SRGD Publications, Department of Computer Science, University of Calgary. Proc-1, pp1-20.
- [2] Castillo, E., Gutiérrez J.M. and Hadi A.S. (1997). *Expert Systems and probabilistic network models*. Springer-Verlag.
- [3] Fagin, R. and Halpern, J.Y. (1994). Reasoning about knowledge and probability. *Journal of the ACM*, 41(2):340–367.
- [4] Fagin, R, Halpern, J.Y. and Megiddo N. (1990). A logic for reasoning about probabilities. In *Information and Computation*. **87**. 78-128.
- [5] Friedman, N., Linial, M., Nachman, I., Pe'er, D. (1999). *Using Bayesian networks to analyze expression data*.
- [6] Hunter, A. (1996) Paraconsistent Logics. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, Volume 2: Reasoning with Actual and Potential Contradictions.
- [7] Minka, T.P. (2001). *A family of algorithms for approximate Bayesian inference*. PhD thesis. MIT Media Lab.