Improvements on KAMET-to-Bayesian-Network Transformations*

Rafael Peñaloza

Instituto Tecnológico Autónomo de México (ITAM)
Department of Computer Science
Rio Hondo 1, Mexico City, Mexico.
rpenalozan@yahoo.com

Abstract. KAMET is a model-based methodology designed to manage knowledge acquisition from multiple knowledge sources that leads to a graphical model that represents causal relations [2]. In the past, all the inference methods developed for KAMET were rule-based; and thus, produced a loss of the uncertainty information included in the models. In [3] two transformations are presented, which allow the use of Bayesian networks as an inference engine for KAMET models that include probabilistic uncertainty. In this paper we present a simple proof of the equivalence between those transformations in the sense of yielding the same results under equivalent evidence, as well as some improvements made to the KAMET Modelling Language and the definition of the transformations (where an error is corrected).

1 Introduction

Probabilistic networks form a good reasoning tool under probabilistic information. Furthermore, the inference methods developed for these networks guarantee the correctness of the solutions given that the information contained on the models is correct. Thus, probabilistic networks have been successfully applied to many different problems, for example, vision systems [6] and cancer diagnosis [7].

Unfortunately, the construction of these networks is not straightforward. In fact, the problem of creating a correct graphical model is equivalent to that of knowledge acquisition, which is still known as a bottle-neck for Knowledge Based Systems (KBS). And this problem is even worse when knowledge comes from multiple sources.

KAMET[1] (Knowledge Acquisition Methodology) is a methodology designed to manage knowledge acquisition from multiple sources. As a result of this methodology, a graphical model that represents causal relations is obtained. In [3],

* 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.

two methods are presented to transform KAMET models into Bayesian networks. Those methods may be used by Knowledge Engineers (KE) to ensure the correctness of the graphical models representing the causal relations between the variables involved.

The transformations are, then, one solution to two big problems: the lack of probabilistic inference methods in KAMET, and the difficulty of forming a correct graphical model in which well known probabilistic inference methods, such as two phase propagation, may work.

The probabilistic networks appear as an alternative to rule based inference methods [9], which were hitherto the only inference methods developed for KAMET models, despite of the uncertainty included in them.

In order to make this paper as clear as possible, we present a brief description of some important elements of KAMET and the transformations for discrete (finite) variables. We have found an *errata* in the process to generate the probability tables, so we give a deeper (and corrected) explanation of it, as well as further development made in both, KAMET modelling and the transformations.

After that, we present a proof of the equivalence of results given by both transformations under equivalent information. Given that equivalence, one is able to choose the transformations that fits a problem's needs.

2 KAMET

For brevity sake, we will give just a brief description of some KAMET elements that are essential for the transformations. For a thorough explanation of KAMET and all its steps, see [2].

As we said before, a graphical model is the result of applying the KAMET methodology for a specific knowledge-acquisition-task. The graphical models summarize as much knowledge as possible, by means of its conceptual modelling language (CML).

The KAMET CML has three levels of abstraction. The first one corresponds to structural constructors and components. The structural constructors are used to highlight the problem itself; they are: *problem*, *classification*, and *subdivision*. The structural components are used to establish the characteristics and possible solutions of the problem. They are: *symptom*, *examination*, *formula*, *inaccurate*, *value*, *antecedent*, *time*, *process* and *solution*. All the elements in this level may be named using an alpha-numeric label. This facilitates the construction and understanding of models.

We aim to the construction of a Bayesian network, which is only concerned about symptoms, problems, and uncertainty (in KAMET *argot* inaccurate); we will also need the structural component value for the generalization. Thus, we will describe only those elements of the first level, whose symbolic representation is shown in figure 1. This does not mean that the other structural components and constructors are futile; actually, it is easy to find applications where each of them is essential, and may not be ignored. What we recommend is to use

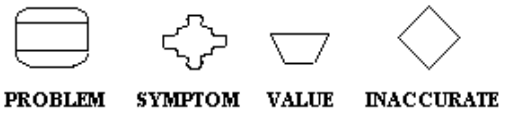

Fig. 1. Symbolic representation of some of KAMET's structural constructors and components.

the Bayesian network for probabilistic inference, but keep the original model as reference for other kinds of knowledge in the domain.

A *symptom* is a manifestation or sign related to an alteration, disorder or abnormality (that is, a *problem*). *Inaccurate* expresses the uncertainty or lack of precision of a node; this uncertainty may be expressed in many different ways, but the most common is using probabilities. The structural component value is always placed above a symptom, to express the characteristics of this symptom. It may include an indicator that will be useful to make restrictions over the possible values.

The second level of abstraction corresponds to nodes and composition rules. Models are represented by acyclic digraphs, where the structural constructors and components form the nodes and *composition rules* (*division*, *implication*, *action* and *union*) form the edges.

The third level of abstraction corresponds to the global model. It should represent the knowledge acquired from multiple knowledge sources in a specific knowledge domain.

The elements of the CML may not be used randomly to form a model. They must follow some diagrammatic conventions and postulates, intended to give more formality and uniformity to the models. We present some of them.

2.1 Diagrammatic Conventions and Postulates

1. Structural components that have times and/or values in common, may be linked into a *group*. Groups are recursive in the sense that they may be elements of other groups. For the purposes of inference, groups act as conjunctions.
2. An *indicator* is represented with a square located in the upper right corner of a group or the structural component value being referred to. It must have one of three kinds of labels: a n is used to express the exact number of elements that must be present; a $n+$ expresses that at least n elements must be present; the n, m label shows the minimum and maximum values that must be present, respectively.
3. An initial node must be a symptom, antecedent, or group. A terminal node represents a structural constructor.
4. Initial and intermediate nodes can be grouped together, without losing their properties or functions, into *molecular nodes*. These will act as nodes in their own right. Molecular nodes are formed through conjunctions or disjunctions.

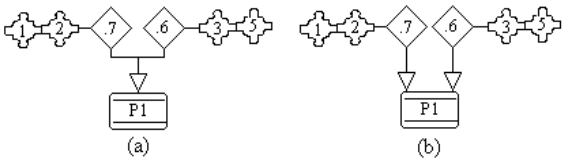

Fig. 2. Disjunctions: (a) a disjunctive molecular node (informative disjunction), and (b) an uninformative disjunction.

5. An *uninformative disjuction* is a disjunction where the model has no information of the uncertainty measure present if more than one of the elements of the disjunction are present. It is represented using many implications (one from each element in the disjunction). Under this definition, a disjunctive molecular node may also be called *informative disjunction* (see figure 2).

3 Transformations

The so called transformations are methods to create Bayesian networks, using the information contained in a KAMET model. As we said before, Bayesian networks are not concerned about many elements of KAMET CML, such as formulas or examinations. Thus, the transformations will work exclusively over the problems and symptoms, using the structural component inaccurate to construct the conditional probability tables. If a node is not binary, then it will have a structural component value associated to it, with an indicator showing the restrictions over the possible values.

It may seem that both structures, Bayesian networks and KAMET models, are incompatible, for the conditional structure of KAMET models is from symptom to problem, while in Bayesian networks it is the complete opposite. By an analysis of the KAMET methodology, it becomes evident that the same process may be applied to make models with a Bayesian-network-compatible conditional structure.

In order to fulfill KAMET postulates, one may use the same kind of graphical representation, with just a reinterpretation of symptom- and problem-nodes. From now on, we will suppose that this modified KAMET methodology was applied, and thus is compatible with probabilistic networks. We will use the same original name of the nodes, but interpret them in the correct fashion.

The construction of the probabilistic models may be divided in two steps: the construction of the graph representing the causal relations, and the construction of the conditional probability tables.

3.1 Graph Construction

Even though probabilistic networks make no distinctions between nodes, we will use different shapes to distinguish the function of the nodes, for the sake of

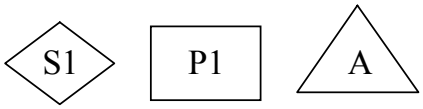

Fig. 3. The three nodes used by the transformations: S1 is a symptom, P1 is a problem and A is an **AND** node.

representation. This will not interfere with the evidence-propagation methods. The three kinds of nodes we will use are shown in figure 3.

We may use any of two transformations, the **AND**-node- and the direct-transformation, to obtain different graphical representations. Each of these graphical models will have advantages and disadvantages respect to the other, as we will show in this section.

The first step of the graph construction consists of creating a node (of the right type) for each symptom and problem that appears in the original KAMET model. Each node will receive the same name it had in the original model, except if it consists of only numeric characters. In that case, we will add a letter at the beginning of the name to avoid confusions with numerical values that the node may take.

For the **AND**-node transformation, we also need to create an artificial binary node, called **AND** node, for each conjunction in the original model. These **AND** nodes will use upper case letters as name, and will have a value of "true" if and only if the conjunction formed by their parents is fulfilled. Note that a disjunction (of any kind) may be seen as a multi-valued node; in fact, every KAMET model with a disjunction may be "rewritten" as a disjunction-free model using multi-valued nodes. Thus, we will suppose, without loss of generality, that these new **AND** nodes do not have any disjunction as a parent.

Now, we just need to add the directed edges. In the direct transformation, all we have to do is to add directed edges to a node, say N, from all the nodes that have implications over N, making no distinctions on the way they are connected.

In the case of the **AND**-node transformation, add directed edges from each of the nodes in a conjunction to the corresponding **AND** node, and then from it to the node implied by the conjunction in the original model. The rest of the edges are like in the direct transformation. Figure 4 shows the result of applying both transformations to the KAMET model in figure 2 (a).

The **AND**-node transformation has the advantage of keeping the graphical representation of KAMET models, showing not just which nodes are related, but which kind of relation they have. Nonetheless, this advantage becomes irrelevant if we decide to keep the original model as we said before.

The "obvious" disadvantage of this transformation is the size of the models it creates, compared to the direct transformation. This is not always a disadvantage, for after the moralization process the **AND**-node models tend to have

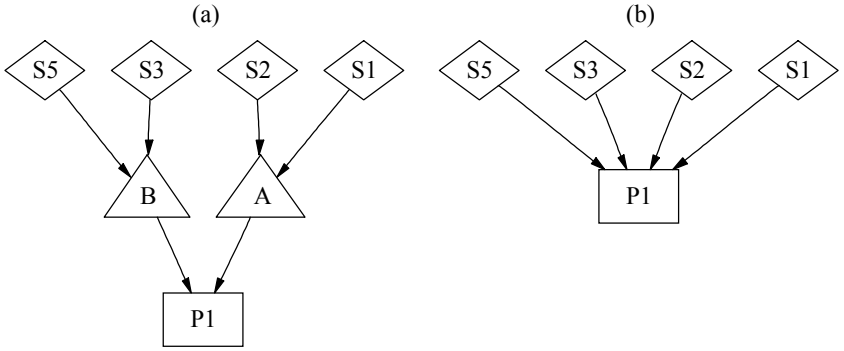

Fig. 4. The resulting graphs of applying the **AND**-node- (a), and the direct-transformation (b) to the KAMET model in figure 2 (a).

many small cliques, while the directly generated ones have few big cliques. Thus, the two phase propagation may work more efficiently over the **AND**-node models, specially if it is done in parallel.

3.2 Table Construction

The table construction presented in [3] is inconsistent with the knowledge contained in the model, as we show next.[2]

Let V be the set of nodes in a probabilistic models obtained by any of the transformations. For each $v \in V$, let C be the set of all possible configurations of the values of the parents of v. For clarity, we will denote the set of parents of v as $pa\,(v)$ and $p\,(v|pa\,(v) = c)$ as $p\,(v|c)$. Let $K = \{c \in C : p\,(v|c)$ is known$\}$, and $K^c := V \backslash K$.

In [3] the construction of conditional probability tables used the probability given the configurations in K, and filled all the unknown information with the prior (known) probability of v. This process produces an inconsistency of information, for the total probability law says

$$p\,(v) = \sum_{c \in C} p\,(v|c)\,p\,(c) = \sum_{c \in K} p\,(v|c)\,p\,(c) + \sum_{c \in K^c} p\,(v|c)\,p\,(c)\,,$$

so, if we use the prior of v to fill in the unknown conditionals, we get that

$$p\,(v) = \sum_{c \in K} p\,(v|c)\,p\,(c) + p\,(v) \sum_{c \in K^c} p\,(c)$$

and this is only true if $\sum_{c \in K} p\,(v|c)\,p\,(c) = p\,(v)\left(1 - \sum_{c \in K^c} p\,(c)\right)$. Thus, it is not difficult to have a model where this does not happen, leading to a contradiction within the knowledge in the model.

[2] We would like to thank Alberto Tubilla for pointing out this *errata*.

Nonetheless, knowledge-consistent approximations may still be made. First suppose that we get to know $p(c)$ for every c in C. Then, we can decide to use the same value (say α) for all the unknown conditionals, for we have no information that allows us to give different values. We then have that

$$p(v) = \sum_{c \in K} p(v|c)\, p(c) + \alpha \sum_{c \in K^c} p(c).$$

But in this equation all the terms, except α, are known; thus, we can conclude that

$$\alpha = \frac{p(v) - \sum_{c \in K} p(v|c)\, p(c)}{\sum_{c \in K^c} p(c)}.$$

If the denominator is zero, then one may give any value (for example 0) to α.

Now, we just have to know $p(c)$. These may be easily approximated using the conditional independence properties of Bayesian networks (which exist because the joint density may be factorized as $p = \prod_{v \in V} p(v|pa(v))$) and the approximations of the conditionals of each node given its parents. This means that, in order to fill all the conditional probability tables, one must start with the initial nodes (for which we have all the information), and approximate the conditionals of their sons, which will be used to approximate their joint distribution, and continue this process through all the "layers" of the graph until all the terminal nodes have their conditional probability tables complete.

Here, the **AND**-node transformation has the disadvantage that adds layers to the graph, increasing the number of operations needed for the approximation process. So, big problems with many conjunctions may need a big amount of additional time, compared to the direct transformation, for this process.

We used the same value for all the unknown conditionals of a given node to fulfill Laplace's principle of insufficient reason, which is a particular case of Maximum Entropy [8]. Of course, other approximations may be done, especially if more information is available. For example, if two of the parents of a node are known to be mutually exclusive, then filling the conditional probability tables with zeros whenever both parents appear may be useful during the propagation stage, while making inferences in the Bayesian networks.

But we have not yet described the set K for a given node $v \in V$. We will describe this set first for **AND**-node models, and then for the other models.

If v is an **AND** node, then K is the set of all the configurations of its parents, because all the conditional probabilities are described for this kind of nodes. If v is not an **AND**-node then, by construction of the models, all of its parents are connected by some kind of disjunction, and we can suppose that they are all **AND** nodes. Whenever this disjunction is informative, then, by definition, all the conditional probabilities are known. If we have an uninformative disjunction, then we just know the conditional probabilities when one (and only one) of them has a value of "true". All the other configurations are part of K^c.

In the case of directly-generated models, the set K may be described basically in the same way, but one must be careful in order to avoid considering some configurations for which the conditional probability is known as unknown. For

example, the model in figure 2 (b) says that if symptoms 1 and 2 are present, then we have problem P1 with probability 0.7, and if the symptoms are 3 and 5, the probability is 0.6. But, what happens if we find out symptoms 1,2 and 3, but not symptom 5? Is this configuration in K?

One may think it is not, for the disjunction is uninformative. Nonetheless, the model says that if symptoms 1 and 2 are present (as they are in this case), then the probability of occurrence of P1 is 0.7. Thus, the configuration is in K and the associated conditional probability must be 0.7.

3.3 Equivalence

Now we want to prove that, if the same evidence is inserted in both models, they will come up with the same results after using exact propagation methods.

Let A be the set of **AND** nodes created for the first transformation, and B the set of the nodes in the original KAMET model, which is equal to the set of nodes in the directly-created Bayesian network. Then, as the information inserted to both models must be the same, in order to prove our claim it is enough to show that the joint probability densities of both models over the set B are equal. If we denote as p_d and p_a the probability density over the direct model and the **AND**-node model, respectively, then we want to show that $p_d = p_{a(B)}$.[3]

From now on, for every node $v \in B$, we will denote as $po(v)$ the set of nodes that are parents of v in the original KAMET model, and $a(v)$ the parents of v in the model with **AND** nodes. We can suppose, without loss of generality, that $a(v) \subset A$.

First, it is important to notice that given a node $v \in B$, every configuration c_p of $pa(v)$ defines one, and only one, configuration c_a of $a(v)$, such that $p_a(a(v) = c_a | po(v) = c_p) = 1$ and $p_a(a(v) = c | po(v) = c_p) = 0$, $\forall c \neq c_a$. Thus,

$$\begin{aligned} p_a(v|po(v) = c_p) &= p_a(v|a(v) = c_a, po(v) = c_p)\, p_a(a(v) = c_a|po(v) = c_p) \\ &= p_a(v|a(v) = c_a)\, p_a(a(v) = c_a|po(v) = c_p) \quad (1) \\ &= p_a(v|a(v) = c_a) \quad (2) \\ &= p_d(v|po(v) = c_p), \quad (3) \end{aligned}$$

where the equality in (1) is given by the independence properties of Bayesian networks, (2) is given by the definition of c_p and c_a and (3) is a consequence of the way the conditional probability tables were constructed.

These equations show that in order to calculate a marginal for a node, we need only the variables in the set B, and that the equations to make such calculus are the same for both models. Then, whenever we get evidence of the state of a collection of variables that are all elements of B, the posterior joint density given that evidence will be the same for both models. This shows that, under exact propagation methods and the same evidence, the two probabilistic models created using each of the transformations yield to the same results.

[3] $p_{a(B)}$ denotes the marginal of the joint density p_a over the set B.

We have shown that, if the same evidence in both models, they will yield the same results. The **AND**-node models are capable of adding extra evidence that will be very difficult in the direct models, however.

For example, suppose that we have the two models shown in figure 4, then in (a) we are capable of inserting the evidence that the node A didn't happen (that is, has a value of "false"). To insert that evidence in the other model, we would have to say that one, but just one, of the symptoms 1 and 2 happened, or none of them. Clearly, that kind of information is easier to insert on **AND**-node models.

Depending on the specific application being developed, one of the models may be a better than the other. The choice of the transformation used is up to the KE. Another option is to use both transformations and, at the end of the process, decide which will fit better the characteristics of the KBS. We don't recommend this option, though, because it means a waste of resources producing models.

4 Conclusions

In this paper we have presented a corrected and expanded description of two KAMET-to-Bayesian-network transformations. We also proved an equivalence between the two probabilistic networks created by the transformations.

The principal purpose of those transformations was the development of a probabilistic inference engine for KAMET, as an alternative to decision rule methods.

Their construction brought an additional advantage: they describe a new method for the construction of causal probabilistic networks, that ensures the correctness of the causal relations in the models obtained, even if the knowledge comes from multiple sources. Thus, the KAMET methodology should be considered by KE as a plausible modelling tool, when planning to use Bayesian networks in a KBS.

KAMET's scopes do not limit to these applications. Other uncertainty measures different from probability may be used, with the help of new inference methods that may be developed in the future, or with the decision rule methods that have been already developed. Even more, the KAMET CML allows the representation of more robust models, that include not only the causes and effects included in Bayesian networks, but also the examinations that may be made to find out the problems, or their possible solutions.

In the process of creation of the transformations, we made some improvements on the CML of KAMET. Further possible improvements to KAMET include the capacity of using variables with an infinity of possible variables, or even continuous random variables, instead of only finite as it does now. This could help to take advantage of more inference methods developed for Bayesian networks.

Current research on the transformations is focused on methods to obtain better approximations for the unknown conditional probabilities. This may be done using the maximum entropy principle.

We are also working on a computational implementation of the transformations and some Bayesian-network inference methods. The goal is to create a tool that will take a KAMET model with probabilistic uncertainty. It will construct a Bayesian network with the transformations specified by the user, and then make some probabilistic inferences based on the probabilistic model.

References

1. Cairó, O. (1997). The KAMET Methodology: A Modelling Approach for Knowledge Acquisition. In Smith and Niku-Lari (Eds.). *Technology Transfer Series*, UK.
2. Cairó, O. (1998). The KAMET Methodology: Content, Usage and Knowledge Modelling. 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, pp. 1-20.
3. Cairó, O. and Peñaloza R. (2003). Using Bayesian Networks as an Inference Engine in KAMET. In *Proceedings of the XXIII International Conference of the Chilean Computer Science Society (SCCC-2003)*. IEEE Press. pp. 79-85.
4. van der Gaag, L.C., and Helsper, E.M. (2002). Experiences with Modelling Issues in Building Probabilistic Networks. In Gómez-Pérez, A., and Benjamins V.R. (Eds.). *EKAW 2002*, LNAI 2473, pp. 21-26.
5. Grünwald, P. (2000). Maximum Entropy and the Glasses You are Looking Through. In *Proceedings of Uncertainty in Artificial Intelligence (UAI-00)*.
6. Jensen, F. V., Christensen, H. I., and Nielsen, J. (1992). Bayesian Methods for Interpretation and Control in Multi-Agent Vision Systems. In *Proceedings: Applications of AI X: Machine Vision and Robotics*.
7. Kahn, C., Roberts, L.M., Wang, K., Jenks, D., and Haddawy, P., (1995). Preliminary Investigation of a Bayesian Network for Mammographic Diagnosis of Breast Cancer. In *Proceedings: American Medical Informatics Association Conference*. pp. 208-212.
8. Perttu, S. (2001). *Uses of Maximum Entropy.* http://www.cs.helsinki.fi/u/myllymak/Teaching/2001/Spring/Seminar/
9. Wong, S.K.M., and Butz, C.J. (1998). *A Probabilistic Network Versus Decision Rules*. http://www2.cs.uregina.ca/~butz/publications/publications.html