Efficient Multiple-Disorder Diagnosis by Strategic Focusing

Linda van der Gaag & Maria Wessels
Utrecht University, Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
e-mail: {linda,mlw}@cs.ruu.nl

Abstract

The belief network framework is becoming increasingly popular for building diagnostic knowledge-based systems. The framework is especially suited for the task of diagnosis because it provides for modelling and dealing with multiple interacting disorders. However, this ability often is exploited insufficiently due to the computational complexity involved. In this paper, we present a method for multiple-disorder diagnosis with a belief network that derives its efficiency from focusing on small sets of related disorders which are constructed by taking advantage of the independencies portrayed by the graphical part of the network.

1 Introduction

Although diagnosing multiple disorders has been a long-standing concern of knowledge-based systems research, it is only recently that fundamental paradigms for dealing with multiple disorders have begun to arise. The paradigms of *model-based reasoning* [Reiter, 1987] and *abductive reasoning* [Peng and Reggia, 1990] especially are tuned to multiple-disorder diagnosis. The techniques developed around these paradigms allow for diagnosing multiple disorders that occur simultaneously, yet do not interact. For example, GDE as an instance of model-based reasoning [de Kleer and Williams, 1987] has been applied most successfully to technical domains where interaction effects of disorders are only occasional. Interaction effects of disorders, however, are very common in for example medical domains. The basic techniques used for model-based reasoning and abductive reasoning lack expressive power for modelling and dealing with such interaction effects.

At present, more and more diagnostic knowledge-based systems are being built using the *belief network framework* [Shwe *et al.*, 1991; Heckerman *et al.*, 1992]. This framework also allows for dealing with multiple disorders. The powerful formalism of the framework provides for explicitly modelling knowledge concerning interaction effects of simultaneous disorders. In addition, the framework has a firm foundation in probability theory and provides for exact and mathematically sound reasoning with uncertain information. This property contrasts approaches for dealing with uncertainty generally employed with model-based reasoning [de Kleer, 1990] and abductive reasoning [Wu, 1990], which often depart from (over-)simplifying assumptions.

The ability of the belief network framework to deal with multiple disorders often is exploited insufficiently since the methods for diagnostic reasoning in use with the framework are over-restrictive. For example, the methods for selective evidence gathering employed aim at gathering information to distinguish between mutually exclusive disorders [Heckerman *et al.*, 1992; van der Gaag and Wessels, 1993]. The often made single-disorder assumption has its origin in the computational complexity involved and is not inherent to the belief network framework itself.

In this paper, we present a method for diagnostic reasoning about multiple, simultaneous and interacting, disorders with a belief network. In Section 2 we review the belief network framework and briefly outline diagnostic reasoning with the framework. Section 3 introduces the basic idea of our method. Section 4 presents a clustering algorithm that lies at the heart of our method which is further detailed in Section 5. The paper is rounded off with some conclusions in Section 6.

2 Preliminaries

The belief network framework provides a formalism for representing a joint probability distribution on a domain of application. A belief network comprises two parts: a qualitative representation and a quantitative representation of the distribution.

The qualitative part of a belief network is a graphical representation of the independencies between the statistical variables discerned in the problem domain; it takes the form of an acyclic digraph $G = (V(G), A(G))$ with nodes $V(G)$ and arcs $A(G)$. Each node in the digraph represents a variable that can take one of a set of values. In the sequel, we will restrict the discussion to binary variables taking one of the values *true* and *false*; the generalization to variables with more than two discrete values, however, is straightforward. We will adhere to the following notational convention: v_i denotes the proposition that the variable V_i takes the value *true*; $V_i = false$ will be denoted as $\neg v_i$. The arcs of the digraph represent dependencies between the variables. Informally speaking, we take an arc $V_i \to V_j$ in the digraph to represent a direct 'influential' relationship between the linked variables, where the direction of the arc designates V_j as the effect of V_i. Absence of an arc between two nodes means that the corresponding variables do not influence each other directly.

The following definitions state the probabilistic meaning of the topology of the digraph of a belief network more formally [Pearl, 1988].

Definition 2.1 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let t be a trail in G. We say that t is* blocked *by a set $W \subseteq V(G)$ if t contains three consecutive nodes X_1, X_2, X_3 for which one of the following conditions holds:*

- *$X_1 \leftarrow X_2$ and $X_2 \to X_3$ are on the trail and $X_2 \in W$;*
- *$X_1 \to X_2$ and $X_2 \to X_3$ are on the trail and $X_2 \in W$;*
- *$X_1 \to X_2$ and $X_2 \leftarrow X_3$ are on the trail, and $\sigma^*(X_2) \cap W = \varnothing$, where $\sigma^*(X_2)$ is the set of nodes composed of X_2 itself and all its descendants.*

The trail t is said to be active *given the set W if it is not blocked by W.*

Building on the notion of blocking, we define the d-separation criterion.

Definition 2.2 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let $X, Y, Z \subseteq V(G)$. The set Y is said to* d-separate *X and Z, denoted as $\langle X \mid Y \mid Z \rangle_G^d$, if for each $V_i \in X$ and $V_j \in Z$ every trail from V_i to V_j in G is blocked by Y.*

The d-separation criterion provides for reading independency statements from a digraph G: if we have $\langle X \mid Y \mid Z \rangle_G^d$, then X and Z are taken to be *conditionally independent* given Y.

Associated with the digraph of a belief network is a numerical assessment of the 'strengths' of the represented relationships: with each node is associated a set of (conditional) probabilities describing the influence of the values of the predecessors of the node on the probabilities of the values of the node itself. These sets provide all information necessary for uniquely defining a joint probability distribution Pr that respects the independency relation portrayed by the digraph of the network. For making probabilistic statements concerning the variables discerned, an inference method is associated with the belief network formalism [Pearl, 1988; Lauritzen and Spiegelhalter, 1988].

The belief network framework is often employed for building diagnostic knowledge-based systems. In diagnostic reasoning, the main objective is to identify the most probable set of disorders that explains the manifestations observed in a specific problem. In the sequel, we will refer to this set as the *diagnosis* of the problem. Establishing a diagnosis is generally achieved by gathering information on the problem at hand. In most domains, it is not necessary nor desirable to obtain information on all possible manifestations before a diagnosis is reached. Hence, diagnostic reasoning is a *strategic* reasoning process in which a selective approach to evidence gathering is taken. The belief network framework does not provide for strategic control over reasoning and therefore is enhanced to this end. Several methods for diagnostic reasoning with a belief network, and selective evidence gathering more in specific, have been proposed in the literature.

For modelling the different roles the nodes of a belief network have in diagnostic reasoning, generally the following types of node are distinguished [Henrion, 1989]: a *hypothesis node* represents a disorder or hypothesis that may be confirmed or disconfirmed; an *evidence node* represents a variable whose value can be obtained by observation; an *intermediate node* represents a variable not classified in either of the former two groups. From now on, we take the nodes of the digraph G of a belief network to be divided into three sets of nodes: $H(G) = \{H_1, \ldots, H_n\}$, $n \geq 1$, is the set of hypothesis nodes; $E(G) = \{E_1, \ldots, E_m\}$, $m \geq 1$, is the set of evidence nodes; $I(G)$ is the set of intermediate nodes.

In the methods for selective evidence gathering in use with a belief network, generally two simplifying assumptions are made. First, a *myopic* approach to evidence gathering is taken, that is, evidence nodes to acquire information on are selected one by one. It is conceivable that in practical applications a non-myopic approach in which nodes are selected groupwise outperforms any method based on a myopic approach. Naively adopting a non-myopic approach, however, poses serious computational problems. Further research aimed at gaining insight in solving these problems is underway [Heckerman *et al.*, 1993]. Since adopting a myopic approach to evidence gathering is quite common in diagnostic knowledge-based systems, we will equally take this approach in this paper. Secondly, the number of nodes in the belief network representing disorders is often restricted to *one*. This restriction prohibits reasoning about multiple interacting disorders. Relaxing this restriction and applying evidence gathering in view of a set of hypothesis nodes straightforwardly also causes problems from a computational point of view; in addition, such an approach would strive to ascribe a value to all hypothesis nodes discerned, which is conceptually unattractive since a diagnosis generally will not encompass all these nodes. In the sequel, we present an intuitively more appealing and more efficient approach in which the latter restriction is eased.

3 Multiple-Disorder Diagnosis and Belief Networks

In many domains, a diagnosis may be composed of simultaneous and interacting disorders. As mentioned before, diagnostic reasoning is the process of establishing a diagnosis for a given problem. Before a diagnosis is reached, generally several possible disorders are considered and either confirmed or disconfirmed. It often is reasonable to assume that not *all* possible disorders discerned in the domain need be investigated. In fact, many possible disorders are never considered for the problem and therefore will not be confirmed or disconfirmed. From the above observations, we have that in the context of a belief network a diagnosis will generally not involve all hypothesis nodes. We will therefore take a *diagnosis* to be a conjunction of values for a subset of the set of hypothesis nodes of the network.

Diagnostic reasoning starts by gathering and processing information that is readily available. In a medical context, examples are the patient's history and information from physical examination. In the sequel, we will use the phrase *surface evidence* to denote this type of information. For a belief network, the surface evidence obtained is a conjunction of values for a subset of the set of evidence nodes.

Diagnostic reasoning proceeds by selecting likely disorders to be considered for the problem at hand. Generally, only disorders that are suggested by the surface evidence are selected at this stage. In a belief network, we have to provide for identifying likely, possibly interacting disorders. To this end, we *partition* the set of hypothesis nodes into disjunct subsets, called *blocks*. These blocks are constructed to be mutually independent given the evidence obtained so far. Hypothesis nodes from different blocks are independent, meaning that the represented disorders are deemed unrelated in the problem under consideration at this stage of the reasoning process. Hypothesis nodes from the same block, on the other hand, may be dependent, meaning that the represented disorders may be related in the problem and have interaction effects. Especially in larger belief networks modelling many, groupwise unrelated disorders, these blocks are expected to be rather small, comprising a few hypothesis nodes each. Upon investigation, a block of hypothesis nodes may contribute a *partial diagnosis* composed of simultaneous, possibly interacting, disorders to the final diagnosis of the problem at hand; separate blocks may yield unrelated partial diagnoses.

The likely disorders that are selected for the problem at hand based on the evidence obtained so far are further investigated by gathering additional information; we will refer to this type of information as *deep evidence*. As opposed to surface evidence, deep evidence generally is not readily available and may for example involve high costs of obtaining. The information to be acquired at this stage of the reasoning process therefore is selected carefully. In a belief network, the evidence nodes related to the likely disorders suggested by earlier acquired evidence will generally be most advantageous to acquire information on. This observation has motivated *clustering* the evidence nodes with respect to the constructed blocks of hypothesis nodes. We then focus selective evidence gathering on the separate blocks of hypothesis nodes and their related evidence nodes.

Newly acquired evidence may contribute to the (dis)confirmation of some of the disorders under consideration. The evidence may also suggest other disorders to be investigated. In addition, it may provide a basis for concluding that in the problem at hand certain hypotheses have become (un)related. The set of likely disorders to be considered thus changes dynamically as further evidence becomes available.

4 Clustering

In the previous section, we have outlined the basic idea of our method for diagnostic reasoning about multiple disorders with a belief network. In this section, we present the concepts involved in our method in general terms; we will turn to their use in diagnostic reasoning in Section 5. In Section 4.1, we introduce the notions of partition and clustering, and state some properties. Section 4.2 presents an algorithm for computing a clustering of the nodes of the digraph of a belief network; the algorithm is illustrated with an example in Section 4.3.

4.1 Definitions and Properties

The clustering of the set of nodes of the digraph of a given belief network that will be exploited in the sequel is built on a partition of the set of hypothesis nodes of this network.

Definition 4.1 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let $X, Y \subseteq V(G)$. A* partition *of X given Y, denoted as $\psi(X \mid Y)$, is a set of disjunct subsets $X_i \subseteq X$, $i = 1, \ldots, p$, $p \geq 1$, called* blocks*, such that*

- $\bigcup_{i=1,\ldots,p} X_i = X$;
- *for all $j, k = 1, \ldots, p$, $j \neq k$, $\langle X_j \mid Y \mid X_k \rangle_G^d$.*

For a given set of nodes Y, each set of nodes X allows at least one partition given Y since $\psi(X \mid Y) = \{X\}$ is such a partition. We introduce the notion of an optimal partition as a partition of which the separate blocks cannot be partitioned any further.

Definition 4.2 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let $X, Y \subseteq V(G)$. A partition $\psi(X \mid Y)$ of X given Y is* optimal *if each block $X_i \in \psi(X \mid Y)$ allows one partition only.*

Lemma 4.3 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let $X, Y \subseteq V(G)$. Then, there is one and only one optimal partition of X given Y.*

Proof. Consider a node $V_i \in X$ and the set $\delta(\{V_i\}) = \{V_k \in X \mid \neg\langle\{V_i\}|Y|\{V_k\}\rangle_G^d\}$, that is, the set of all nodes from X that are not d-separated from V_i by Y; note that $V_i \in \delta(\{V_i\})$. It will be evident that this set of nodes is determined uniquely by the set Y and the topology of the digraph G. For any partition $\psi(X|Y)$ of X given Y we have that there exists a block $X_j \in \psi(X|Y)$ such that $\delta(\{V_i\}) \subseteq X_j$. Applying the argument recursively, we find that there exists a block $X_j \in \psi(X|Y)$ such that $\delta^*(\{V_i\}) \subseteq X_j$, where δ^* denotes the reflexive and transitive closure of δ. Now, let $\hat{\psi}(X|Y)$ be an optimal partition of X given Y and let $X_j \in \hat{\psi}(X|Y)$ be the block with $\delta^*(\{V_i\}) \subseteq X_j$. From $\hat{\psi}(X|Y)$ being optimal, we have that $X_j = \delta^*(\{V_i\})$. Then, for each block $X'_j \in \hat{\psi}(X|Y)$, we have that there is a vertex $V'_i \in X$ such that $X'_j = \delta^*(\{V'_i\})$. Since for each vertex V'_i the set $\delta^*(\{V'_i\})$ is unique, we conclude that there is one optimal partition of X given Y only. □

An independency relation may change as evidence becomes available. New independencies may arise; yet, it is also possible that current independencies will no longer hold after observing the evidence. In the sequel, we want to distinguish between independencies that hold no matter which evidence may become available and independencies that may be invalidated by new evidence. To this end, we distinguish between strong and weak d-separation.

Definition 4.4 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let $X, Y, Z \subseteq V(G)$. The sets X and Z are said to be* strongly d-separated *by the set Y if $\langle X|Y|Z\rangle_G^d$ and $\langle X|Y \cup W|Z\rangle_G^d$ for all $W \subseteq V(G)$. The sets X and Z are said to be* weakly d-separated *by Y if $\langle X|Y|Z\rangle_G^d$ and X and Z are not strongly d-separated by Y.*

Since an independency relation may change as new evidence becomes available, the optimal partition of a given set of nodes X may change as well. Observe that for any partition of X we have that any two nodes from different blocks are either weakly d-separated or strongly d-separated. For an optimal partition, we have the additional property that two nodes from the same block are either weakly d-separated or not d-separated at all. From these observations, it follows that the changes to the optimal partition occasioned by a new piece of evidence are restricted in scope.

Lemma 4.5 *Let $G = (V(G), A(G))$ be an acyclic digraph and let $X, Y \subseteq V(G)$. Let $\psi(X|Y)$ be the optimal partition of X given Y. Now, let $V \in V(G)$ and let $Y' = Y \cup \{V\}$. Then, the optimal partition $\psi'(X|Y')$ of X given Y' satisfies one and only one of the following conditions:*

- *$\psi'(X|Y') = \psi(X|Y)$;*
- *$\psi(X|Y) = \{X_1, \ldots, X_p\}$, $p \geq 1$, and $\psi'(X|Y') = (\psi(X|Y) \setminus \{X_i\}) \cup \{X_{i_1}, \ldots, X_{i_q}\}$, $q > 1$, such that $\bigcup_{j=1,\ldots,q} X_{i_j} = X_i$, for some $1 \leq i \leq p$;*
- *$\psi'(X|Y') = \{X_1, \ldots, X_p\}$, $p \geq 1$, and $\psi(X|Y) = (\psi'(X|Y') \setminus \{X_j\}) \cup \{X_{j_1}, \ldots, X_{j_q}\}$, $q > 1$, such that $\bigcup_{i=1,\ldots,q} X_{j_i} = X_j$, for some $1 \leq j \leq p$.*

From the lemma we have that if a new piece of evidence occasions a change to the optimal partition, then either one block of the current partition falls apart into two or more smaller blocks, or two or more blocks of the current partition are combined into one large block.

The optimal partition of a set of nodes X given a set of nodes Y includes only the nodes from X of the digraph at hand. This optimal partition is now extended to a clustering by inserting as many of the remaining nodes into the separate blocks of the partition as possible, exploiting the d-separation criterion.

Definition 4.6 *Let $G = (V(G), A(G))$ be an acyclic digraph and let $X, Y \subseteq V(G)$. Let $\psi(X|Y) = \{X_1, \ldots, X_p\}$, $p \geq 1$, be the optimal partition of X given Y. For $i = 1, \ldots, p$, the* cluster *S_i associated with block X_i of $\psi(X|Y)$ is the set of nodes $S_i = \{V \in V(G) | \neg\langle X_i|Y|\{V\}\rangle_G^d\}$. The* clustering *of $V(G)$ with respect to X given Y, denoted as $\zeta(X|Y)$, is the set $\zeta(X|Y) = \{S_1, \ldots, S_p\}$ where S_i is the cluster associated with block X_i of $\psi(X|Y)$, $i = 1, \ldots, p$.*

Note that the clustering $\zeta(X \mid Y)$ of the set of nodes of a digraph G with respect to a set of nodes X given a set of nodes Y is unique. Also, note that this clustering in general is not a partition of $V(G)$: different clusters may have nodes in common and there may be nodes that are not included in any cluster.

4.2 The Clustering Algorithm

In this section, we present an algorithm for computing for a digraph G the clustering $\zeta(X \mid Y)$ of the set of nodes $V(G)$ with respect to a set of nodes X given a set of nodes Y. The main procedure of our clustering algorithm is the cluster-nodes procedure. This procedure takes two sets of nodes X and Y for its input and yields the clustering $S = \zeta(X \mid Y)$ as its output. The procedure repeatedly selects a node V_i from the set X and computes the cluster comprising V_i, by calling the dependent-nodes procedure; it collects the thus computed clusters into the clustering S.

```
procedure cluster-nodes(X,Y,S)
  T := X;
  while T ≠ ∅ do
      select V_i ∈ T;
      dependent-nodes(X,V_i,Y,S);
      T := T \ {V_i}
  od
end
```

The dependent-nodes procedure is based to a large extent upon the algorithm by Geiger *et al.* for identifying all (in)dependencies implied by the topology of a digraph [Geiger *et al.*, 1990]. For a node $V_i \in X$, the procedure determines the set of all nodes that can be reached from V_i via an active trail given Y in G. Note that this set not necessarily constitutes the cluster for node V_i as a cluster may also contain nodes that are weakly d-separated. This situation only arises when the set of non-d-separated nodes for a node from X comprises at least one other node from X. In the dependent-nodes procedure, therefore, the set of non-d-separated nodes computed for a node $V_i \in X$ is taken as a *temporary cluster*. For each newly created temporary cluster it is verified whether it has nodes from X in common with earlier computed temporary clusters; if so, the set of temporary clusters is adapted by combining two or more temporary clusters into a new one.

```
procedure dependent-nodes(X,V_i,Y,S)
  T := {V_i};
  for each V_j ∈ V(G) do
      if there is an active trail given Y in G
        from V_i to V_j then
        T := T ∪ {V_j}
      fi
  od;
  while there is a temporary cluster T' ∈ S
        with T' ∩ T ∩ X ≠ ∅ do
      T := T ∪ T';
      S := S \ {T'}
  od;
  S := S ∪ {T}
end
```

The correctness of our algorithm follows from the observation that the algorithm performs the construction of the optimal partition given in the proof of Lemma 4.3 straightforwardly.

To conclude, we note that the worst-case computational complexity of the cluster-nodes procedure is polynomial in the number of nodes of the digraph.

4.3 An Example

Consider the example digraph G depicted in Figure 1. Let X be the set of nodes drawn in shading and let $Y = \varnothing$. The cluster-nodes procedure is called upon to compute the clustering $\zeta(X|Y)$ of $V(G)$ with respect to X given Y.

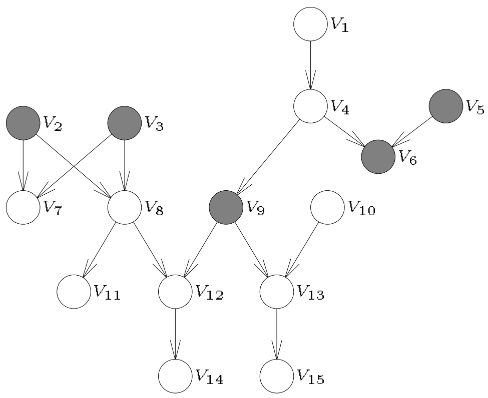

Figure 1: An Example Digraph.

Suppose that node V_5 is the first node selected from X by the cluster-nodes procedure. For this node, the dependent-nodes procedure computes the temporary cluster $T_1 = \{V_5,V_6\}$. This temporary cluster is inserted into the (also temporary) clustering S. Now suppose that subsequently node V_9 is selected from X. For V_9, the temporary cluster $T_2 = \{V_1,V_4,V_6,V_9,V_{12},V_{13},V_{14},V_{15}\}$ is computed. The dependent-nodes procedure detects that node $V_6 \in T_2 \cap X$ has been included earlier in the temporary cluster T_1; T_1 and T_2 therefore are combined into the new temporary cluster T_2, and T_1 is removed from S. The final clustering S contains the three clusters $\{V_2,V_7,V_8,V_{11},V_{12},V_{14}\}$, $\{V_3,V_7,V_8,V_{11},V_{12},V_{14}\}$ and $\{V_1,V_4,V_5,V_6,V_9,V_{12},V_{13},V_{14},V_{15}\}$. Note that in this clustering, node V_{10} does not appear in any of the clusters, and that nodes V_{12} and V_{14} are comprised in all clusters.

If the cluster-nodes procedure had started with node V_6 instead of V_5, then the necessity of combining the temporary clusters T_1 and T_2 would not have arisen. In general, combining temporary clusters is minimized by selecting the nodes from X in decreasing topological order. Combining temporary clusters, however, cannot be avoided in all.

5 The Method for Multiple-Disorder Diagnosis

The main procedure of our method for diagnostic reasoning about multiple, simultaneous and interacting, disorders with a belief network is the multiple-disorder-diagnosis procedure. This procedure takes the digraph G of a belief network and the set E of nodes corresponding with deep evidence for its input and yields a diagnosis D as its output.

procedure multiple-disorder-diagnosis(G,E,D)
  enough := *false*;
  while $E \neq \varnothing$ **and not** enough **do**
    cluster-nodes($H(G)$,$E(G) \setminus E$,S);
    select-cluster(S,S_i,enough);
    if enough = *false* **then**
      select-node(S_i,E,E_j);
      process-evidence(E_j);
      $E := E \setminus \{E_j\}$
    fi

```
  od;
  diagnosis(D)
end
```

The algorithm begins by computing the clustering of the set of nodes $V(G)$ with respect to the set of hypothesis nodes $H(G)$ given the set of nodes for which evidence has been obtained (initially, the set of nodes modelling surface evidence). To this end, the cluster-nodes procedure from the previous section is used as will be further detailed in Section 5.1.

From the clustering yielded by the cluster-nodes procedure, one cluster is selected to focus attention on by the select-cluster procedure, which is described in Section 5.2. From the selected cluster, the evidence node to best acquire information on is determined by the select-node procedure, detailed in Section 5.3. The user then is prompted for a value for the selected node and the value entered is processed in the network by the process-evidence procedure.

Before the next evidence node is selected, it is verified whether or not enough evidence has been gathered to establish a diagnosis. Further evidence gathering may be stopped as soon as for all relevant clusters a partial diagnosis has been found. To verify whether information gathering may be stopped, a *local* stopping criterion for examining sufficiency of confirmation per cluster is employed as well as a *global* stopping criterion for examining relevancy of clusters. In Section 5.2, we briefly describe the latter stopping criterion; here we will not further elaborate on the local stopping criterion. After evidence gathering has completed, the diagnosis procedure collects all partial diagnoses that have been confirmed to sufficient extent and presents these to the user.

At several stages of the reasoning process, strategic decisions are taken as to the focus of attention. These decisions are based on probabilistic considerations. In Section 5.4 an example illustrates how the probabilities involved are computed efficiently.

5.1 Computing the Clustering

Our method for multiple-disorder diagnosis focuses attention on mutually independent subsets of hypothesis nodes. To this end, the set of hypothesis nodes is partitioned into blocks as described before. These blocks should be as small as possible to allow for controlling the computational complexity involved. By departing from the optimal partition of the set of hypothesis nodes the idea of focusing is exploited to the fullest extent since for this partition the blocks are smallest.

The cluster-nodes procedure from the previous section is used to compute the clustering of the set of nodes $V(G)$ with respect to the set $H(G)$ given the set of nodes $\bar{E}$ for which evidence has been obtained. Now note that, instead of just the optimal partition of $H(G)$ given $\bar{E}$, this procedure computes the sets of dependent evidence and intermediate nodes along with the blocks of hypothesis nodes. We have chosen to compute these sets for all blocks of hypothesis nodes along with the partition, since the set of dependent evidence nodes needs be computed anyway for the subset of hypothesis nodes on which attention will be focused: for this subset, a node to acquire information on needs be selected among the dependent evidence nodes. Also, the sets of dependent evidence and intermediate nodes can be yielded as a byproduct of computing the optimal partition without adding to the time complexity of the computation.

After a new piece of evidence has been processed in the belief network, the clustering of the set of nodes of the digraph is computed all over again. From Lemma 4.5, however, we have that the optimal partition does not change drastically for a new piece of evidence. We expect that a more dynamic approach to computing the new clustering is possible by maintaining the current clustering and only re-computing clusters for the part of the network influenced by the new evidence.

5.2 Selecting a Cluster

Consider the clustering of the set of nodes of the digraph G given the evidence obtained so far. The basic idea of the select-cluster procedure is to select the cluster that at this stage of the reasoning process is most likely to yield a partial diagnosis of the problem under consideration. This procedure takes as a constraint that the selected cluster is not contradicted by the evidence.

For each cluster S_i of the current clustering, the probability $Pr(\bigvee_{H_j \in S_i} h_j \mid c)$, where c is the conjunction of all evidence obtained so far, expresses the likelihood that at least one of the hypothesis nodes from the cluster has adopted the value *true* in the problem at hand: the higher this probability, the more likely this cluster will provide a partial diagnosis. In addition, we observe that the cluster is not contradicted by the evidence if $Pr(\bigvee_{H_j \in S_i} h_j \mid c) \geq Pr(\bigvee_{H_j \in S_i} h_j)$. The select-cluster procedure therefore selects the cluster S_i with the highest probability $Pr(\bigvee_{H_j \in S_i} h_j \mid c)$ for which $Pr(\bigvee_{H_j \in S_i} h_j \mid c) \geq Pr(\bigvee_{H_j \in S_i} h_j)$. Note that this procedure encompasses a global criterion for stopping further gathering of information: if, after several clusters have been investigated, all remaining clusters are contradicted by the evidence, then evidence gathering is stopped.

Several probabilities are involved in selecting a cluster to focus attention on. For each cluster S_i, the probabilities $Pr(\bigvee_{H_j \in S_i} h_j)$ and $Pr(\bigvee_{H_j \in S_i} h_j \mid c)$ have to be computed, or the probabilities $Pr(\bigwedge_{H_j \in S_i} \neg h_j)$ and $Pr(\bigwedge_{H_j \in S_i} \neg h_j \mid c)$, alternatively. The prior probability $Pr(\bigwedge_{H_j \in S_i} \neg h_j)$ is computed from the joint probability distribution defined by the belief network by marginalization [Suermondt and Cooper, 1991]. The conditional probability $Pr(\bigwedge_{H_j \in S_i} \neg h_j \mid c)$ is computed basically by employing the recursive approach used in loop cutset conditioning [Pearl, 1988]. The fact that the clustering of the set of nodes may change when new evidence becomes available interferes with this recursive approach.

Consider the optimal partition defining the current clustering. From Lemma 4.5 we have that for a new piece of evidence either one block of the current partition falls apart into two or more smaller blocks, or two or more blocks of the current partition are combined into a new one. First, consider the case where the block B of the current optimal partition falls apart into separate blocks $B_1, \ldots, B_q$, $q > 1$. Let c once more denote the conjunction of all evidence obtained so far and let e denote the new piece of evidence occasioning the change in the current partition. The probabilities $Pr(\bigwedge_{H_j \in B_i} \neg h_j \mid c \wedge e)$, $i = 1, \ldots, q$, required for selecting a cluster in the next step of the reasoning process are computed from the probability $Pr(\bigwedge_{H_j \in B} \neg h_j \mid c)$ by further conditioning on e and subsequent marginalization. Now consider the case where several blocks $B_1, \ldots, B_q$, $q > 1$, of the current optimal partition combine into one block B. The required probability $Pr(\bigwedge_{H_j \in B} \neg h_j \mid c \wedge e)$ is computed using the aforementioned recursive approach and exploiting the property that in the current partition $B_1, \ldots, B_q$ are independent given the evidence obtained so far, that is, $Pr(\bigwedge_{H_j \in B} \neg h_j \mid c) = \prod_{i=1,\ldots,q} Pr(\bigwedge_{H_j \in B_i} \neg h_j \mid c)$. We will further elaborate on these observations in Section 5.4.

To conclude, we note that the computation of the prior probabilities is expensive; the worst-case computational complexity is dependent on the number of hypothesis nodes comprised in the network [Suermondt and Cooper, 1991]. These probabilities, however, need only be computed once before diagnostic reasoning is started. All further computations have a worst-case complexity that is exponential in the number of hypothesis nodes comprised in the separate clusters.

5.3 Selecting an Evidence Node

After the cluster that is most likely to yield a partial diagnosis has been selected, the evidence node to best acquire information on is determined. To this end, the select-node procedure applies the method for evidence gathering presently in use for belief networks: for each evidence node that is not yet instantiated, the expected utility with respect to the set of hypothesis nodes comprised in the selected cluster is computed, and the evidence node with the maximum expected utility is selected. Only evidence nodes from the selected cluster are considered as these are the only evidence nodes relevant to the current focus of attention.

For selecting evidence nodes, several utility functions may be employed. An example is the *linear value function* which is built on the notion of confirmation [van der Gaag and Wessels, 1993]. Let B be the block of hypothesis nodes in the selected cluster and let c_B be a conjunction of values for the nodes in B; let c be the conjunction of all evidence obtained so far and let E_j be an evidence node from the selected cluster. Then, $u(c_B, e_j) = |Pr(c_B \mid c \wedge e_j) - Pr(c_B \mid c)|$

select-cluster		select-node		
step 1				
$Pr(\neg v_2)$		$Pr(v_5 \wedge v_6 \wedge v_9)$	$Pr(v_5 \wedge v_6 \wedge v_9 \mid E_j)$	$Pr(E_j)$
$Pr(\neg v_3)$		$Pr(v_5 \wedge v_6 \wedge \neg v_9)$	$Pr(v_5 \wedge v_6 \wedge \neg v_9 \mid E_j)$	
$Pr(\neg v_5 \wedge \neg v_6 \wedge \neg v_9)$		$Pr(v_5 \wedge \neg v_6 \wedge v_9)$	$Pr(v_5 \wedge \neg v_6 \wedge v_9 \mid E_j)$	
		$Pr(v_5 \wedge \neg v_6 \wedge \neg v_9)$	$Pr(v_5 \wedge \neg v_6 \wedge \neg v_9 \mid E_j)$	
		$Pr(\neg v_5 \wedge v_6 \wedge v_9)$	$Pr(\neg v_5 \wedge v_6 \wedge v_9 \mid E_j)$	
		$Pr(\neg v_5 \wedge v_6 \wedge \neg v_9)$	$Pr(\neg v_5 \wedge v_6 \wedge \neg v_9 \mid E_j)$	
		$Pr(\neg v_5 \wedge \neg v_6 \wedge v_9)$	$Pr(\neg v_5 \wedge \neg v_6 \wedge v_9 \mid E_j)$	
		$Pr(\neg v_5 \wedge \neg v_6 \wedge \neg v_9)$	$Pr(\neg v_5 \wedge \neg v_6 \wedge \neg v_9 \mid E_j)$	
block $\{V_5, V_6, V_9\}$ is selected.		evidence node V_4 is selected, $V_4 = true$ is observed.		
step 2				
$Pr(\neg v_2 \mid v_4)$	$Pr(\neg v_2)$	$Pr(v_2 \mid v_4)$	$Pr(v_2 \mid v_4 \wedge E_j)$	$Pr(E_j \mid v_4)$
$Pr(\neg v_3 \mid v_4)$	$Pr(\neg v_3)$	$Pr(\neg v_2 \mid v_4)$	$Pr(\neg v_2 \mid v_4 \wedge E_j)$	
$Pr(\neg v_5 \wedge \neg v_6 \mid v_4)$	$Pr(\neg v_5 \wedge \neg v_6)$			
$Pr(\neg v_9 \mid v_4)$	$Pr(\neg v_9)$			
block $\{V_2\}$ is selected.		evidence node V_{11} is selected, $V_{11} = true$ is observed		
step 3				
$Pr(\neg v_2 \wedge \neg v_3 \mid v_4 \wedge v_{11})$	$Pr(\neg v_2 \wedge \neg v_3)$			
$Pr(\neg v_5 \wedge \neg v_6 \mid v_4 \wedge v_{11})$	$Pr(\neg v_5 \wedge \neg v_6)$			
$Pr(\neg v_9 \mid v_4 \wedge v_{11})$	$Pr(\neg v_9)$			
block $\{V_2, V_3\}$ is selected.				

Table 1: The Probabilities.

indicates the change in confirmation of c_B if evidence e_j is observed for E_j. The expected utility of acquiring information on E_j is computed as the sum of the utilities of all conjunctions of values for the hypothesis nodes in B and all values for E_j weighted with their probabilities.

Note that selecting an evidence node is performed with respect to only a (small) subset of nodes of the digraph. The computational complexity is exponential in the number of hypothesis nodes comprised in the separate clusters.

5.4 An Example

In this section, we illustrate the multiple-disorder-diagnosis procedure and focus on the computation of the probabilities involved. Let the digraph G depicted in Figure 1 be the graphical part of a belief network. The nodes drawn in shading are hypothesis nodes; the nodes V_8, V_{12} and V_{13} are intermediate nodes; the remaining nodes are evidence nodes. For ease of exposition, we assume that all evidence nodes model deep evidence. Table 1 summarizes the probabilities that are required by the select-cluster and select-node procedures.

Initially, three clusters are computed for G as was illustrated in Section 4.3. These clusters comprise the blocks $\{V_2\}$, $\{V_3\}$, and $\{V_5,V_6,V_9\}$ of hypothesis nodes. For these blocks, the select-cluster procedure computes the probabilities shown in the first row of Table 1 by marginalization from the distribution defined by the network. Now suppose that of these the probability $Pr(\neg v_5 \wedge \neg v_6 \wedge \neg v_9)$ is smallest. Then, the cluster containing the block $\{V_5,V_6,V_9\}$ is selected to focus attention on.

The select-node procedure determines the evidence node to best acquire information on by computing the expected utilities of each of the evidence nodes E_j comprised in the selected cluster, that is, of the nodes V_1, V_4, V_{14} and V_{15}. To this end, several prior and posterior probabilities are required as shown in Table 1. The prior probabilities are computed once more by marginalization. The posterior probabilities are computed by employing the recursive approach proposed before. For example, we have that

$$Pr(v_5 \wedge v_6 \wedge v_9 \mid v_1) = \alpha \cdot Pr(v_1 \mid v_5 \wedge v_6 \wedge v_9) \cdot Pr(v_5 \wedge v_6 \wedge v_9)$$

where α is a normalization constant. $Pr(v_1 \mid v_5 \wedge v_6 \wedge v_9)$ is computed directly from the belief network; $Pr(v_5 \wedge v_6 \wedge v_9)$ has been computed before and need not be re-computed. Now, suppose that evidence node V_4 is selected to acquire information on and that the value v_4 is processed in the network. As a result, the block $\{V_5,V_6,V_9\}$ falls apart into the blocks $\{V_5,V_6\}$ and $\{V_9\}$.

In the next step, the select-cluster procedure requires the probabilities shown in the second row of the table. We observe that $Pr(\neg v_2 \mid v_4)$ need not be computed: from the independencies of the digraph of the belief network we have that $Pr(\neg v_2 \mid v_4) = Pr(\neg v_2)$. A similar observation holds for $Pr(\neg v_3 \mid v_4)$. The probabilities not yet known are computed by marginalization from the probabilities found in the previous step of the algorithm. Suppose that the cluster containing the block $\{V_2\}$ is selected to focus attention on.

For determining the node to best acquire information on, the select-node procedure examines the evidence nodes V_7, V_{11}, and V_{14}. Suppose that node V_{11} is selected and that the evidence v_{11} is entered. As a result, the clusters containing the blocks $\{V_2\}$ and $\{V_3\}$ are combined.

In the third step, the select-cluster procedure once more requires several probabilities as shown in the table. We examine the computation of $Pr(\neg v_2 \wedge \neg v_3 \mid v_4 \wedge v_{11})$. We have

$$Pr(\neg v_2 \wedge \neg v_3 \mid v_4 \wedge v_{11}) = \alpha \cdot Pr(v_{11} \mid \neg v_2 \wedge \neg v_3 \wedge v_4) \cdot Pr(\neg v_2 \wedge \neg v_3 \mid v_4)$$

where α once more is a normalization constant. The probability $Pr(v_{11} \mid \neg v_2 \wedge \neg v_3 \wedge v_4)$ is computed directly from the belief network; note that the evidence v_4 has no influence on the probability of v_{11}. The probability $Pr(\neg v_2 \wedge \neg v_3 \mid v_4)$ can be computed by exploiting the independencies of the digraph of the network:

$$Pr(\neg v_2 \wedge \neg v_3 \mid v_4) = Pr(\neg v_2 \mid v_4) \cdot Pr(\neg v_3 \mid v_4) = Pr(\neg v_2) \cdot Pr(\neg v_3)$$

Note that for this computation there is no need to resort to belief network inference as the probabilities $Pr(\neg v_2)$ and $Pr(\neg v_3)$ have been computed before. Now suppose that the cluster containing the block $\{V_2, V_3\}$ is selected and that the probability of one of the conjunctions of values of V_2 and V_3 is considered to be confirmed to sufficient extent. Now, if the posterior probabilities computed for the other blocks surpass their prior probabilities, evidence gathering is stopped. The confirmed conjunction of values of V_2 and V_3 is returned as the diagnosis.

To conclude our example, we observe that the number of probabilities computed in our approach is far less than the number of probabilities a straightforward approach to selective gathering of evidence in view of multiple-disorder diagnosis would require: in such an approach probabilities for all possible conjunctions of values for all hypothesis nodes would have to be computed.

6 Conclusion

In this paper we have presented a method for diagnostic reasoning about multiple, simultaneous and interacting, disorders with a belief network. Our algorithm builds on a decomposition of the set of nodes of the digraph of a belief network into clusters, each comprising a set of related hypothesis nodes and their dependent evidence and intermediate nodes given the evidence obtained so far. This clustering allows for intelligent control over reasoning by focusing attention on subsets of hypothesis nodes. By focusing on a single cluster that is likely to provide a (partial) diagnosis, computations can be performed local to this cluster, thus alleviating the overall computational burden.

Future research will address optimizing the clustering algorithm: a more dynamic approach in which clusters are re-computed only for the part of the network influenced by new evidence is likely to further save on computational effort. In addition, we project performing experiments on different classes of randomly generated belief networks to gain insight into the sizes of the clusters arising during reasoning. To yield accurate insight into the overall behaviour of our method, it will have to be evaluated in real-life applications as well.

References

[de Kleer and Williams, 1987] J. de Kleer and B.C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, vol. 32, pp. 97–130, 1987.

[de Kleer, 1990] J. de Kleer. Using crude probability estimates to guide diagnosis. *Artificial Intelligence*, vol. 45, pp. 381–391, 1990.

[Geiger *et al.*, 1990] D. Geiger, T. Verma, and J. Pearl. *d*-separation: from theorems to algorithms. In M. Henrion, R.D.Shachter, L.N. Kanal, and J.F. Lemmer, eds, *Uncertainty in Artificial Intelligence 5*, pp. 139–148. Elsevier Science Publishers BV, 1990.

[Heckerman *et al.*, 1992] D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani. Toward normative expert systems. Part 1: The Pathfinder project. *Methods of Information in Medicine*, vol. 31, pp. 90–105, 1992.

[Heckerman *et al.*, 1993] D.E. Heckerman, E.J. Horvitz, and B. Middleton. An approximate non-myopic computation for value of information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 292–298, 1993.

[Lauritzen and Spiegelhalter, 1988] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, vol. 50, pp. 157–224, 1988.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, 1988.

[Peng and Reggia, 1990] Y. Peng and J.A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving.* Springer-Verlag, 1990.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, vol. 32, pp. 57–95, 1987.

[Shwe *et al.*, 1991] M.A. Shwe, B. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann, and G.F. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base: The probabilistic model and inference algorithms. *Methods of Information in Medicine*, vol. 30, pp. 241–255, 1991.

[Suermondt and Cooper, 1991] H.J. Suermondt and G.F. Cooper. Initialization for the method of conditioning in Bayesian belief networks. *Artificial Intelligence*, vol. 50, pp. 83–94, 1991.

[van der Gaag and Wessels, 1993] L.C. van der Gaag and M.L. Wessels. Selective evidence gathering for diagnostic belief networks. *AISB Quarterly*, no. 86, pp. 23–34, 1993.

[Wu, 1990] T.D. Wu. Probabilistic evaluation of candidates and symptom clusterings for multidisorder diagnosis. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pp. 460–467, 1990.