ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-693

The Persistence of Most Probable Explanations in Bayesian Networks

Arnoud Pastink[1] and **Linda C. van der Gaag**

Abstract. Monitoring applications of Bayesian networks require computing a sequence of most probable explanations for the observations from a monitored entity at consecutive time steps. Such applications rapidly become impracticable, especially when computations are performed in real time. In this paper, we argue that a sequence of explanations can often be feasibly computed if consecutive time steps share large numbers of observed features. We show more specifically that we can conclude persistence of an explanation at an early stage of propagation. We present an algorithm that exploits this result to forestall unnecessary re-computation of explanations.

1 INTRODUCTION

Bayesian networks are nowadays being applied for a range of problems in the field of biomedicine. Most notably, networks are being used for diagnostic purposes. Such networks typically capture disease processes and tend to include a single diagnostic variable for which posterior distributions are computed. Bayesian networks have so far been little used for monitoring problems in which deviations from expected behaviour are to be identified. We currently are developing such a Bayesian network, for monitoring the welfare and productivity of a pig herd. Our network differs from more standard Bayesian networks for diagnostic problem solving, in the sense that it includes multiple variables of interest for which posterior joint probabilities need to be computed in view of a sequence of dependent observation vectors. Based upon these probabilities, the most probable joint value assignment over the variables of interest needs to be established. Unfortunately, finding most probable explanations for just a single observation vector from a network is already known to be a computationally intensive task in general [3].

In view of the unfavourable runtime complexity of computing a most probable explanation, we address in this paper the computation of consecutive explanations for a sequence of dependent observation vectors. We assume that in a monitoring application at hand, the underlying processes do not vary disruptively; in fact, we assume that two consecutive observation vectors differ in their values for a single variable only. Given a most probable explanation at some time step, we show that although a new observation may change the explanation, the new explanation cannot become any arbitrary value combination. We further identify conditions under which parts of an explanation are guaranteed to persist over time, and show that these conditions can be readily verified locally upon junction-tree propagation. Based upon these considerations, we present a new algorithm for computing most probable explanations that is tailored to monitoring applications. The algorithm decides at a very early stage during propagation if part of the current most probable explanation will persist in view of a new observation and halts propagation at the earliest possible moment. The algorithm thereby effectively forestals unnecessary re-computation of explanations.

The paper is organised as follows. In Section 2 we introduce our notational conventions and briefly review junction trees for Bayesian networks in general. In Section 3 we show that a new observation can change a constituent explanation from a clique only to one of a predetermined set of value combinations. In Section 4, we show under which conditions persistence of constituent explanations is guaranteed. In Section 5, we combine our persistency results with the basic ideas underlying the cautious and max-propagation algorithms to arrive at a new, practical algorithm for computing most probable explanations in monitoring applications. The paper is rounded off with our conclusions and suggestions for future research in Section 6.

2 PRELIMINARIES

We consider a finite set $\mathbf{V}$ of random binary variables; each variable $V_i \in \mathbf{V}$ takes its value from the associated domain $\Omega_{V_i} = \{\bar{v}_i, v_i\}$. The notation $\mathbf{v}$ is used to denote a joint value combination for all variables from $\mathbf{V}$; the set of all such value combinations is $\Omega_{\mathbf{V}} = \times_{V_i \in \mathbf{V}} \Omega_{V_i}$. If a value combination $\mathbf{v} \in \Omega_{\mathbf{V}}$ and a combination $\mathbf{w} \in \Omega_{\mathbf{W}}$ for a subset $\mathbf{W} \subset \mathbf{V}$ assign the same values to their shared variables, we say that $\mathbf{v}$ is consistent with $\mathbf{w}$ and vice versa. For our monitoring context, we further assume that the set $\mathbf{V}$ is partitioned into a set $\mathbf{C}$ of explanatory variables and a set $\mathbf{X}$ of observable variables, with $\mathbf{C} \cap \mathbf{X} = \varnothing$, $\mathbf{C} \cup \mathbf{X} = \mathbf{V}$. A joint value combination $\mathbf{x} \in \Omega_{\mathbf{X}}$ will be called an observation vector. To distinguish between observations at different times, we add a time tag to each observation vector and write $\mathbf{x}^t$ for the vector of observations at time t. Without loss of generality, we assume that two consecutive vectors $\mathbf{x}^t$ and $\mathbf{x}^{t+1}$ differ in their values for just a single variable $X_i \in \mathbf{X}$.

Over the set of variables $\mathbf{V}$, we further consider a joint probability distribution $\Pr(\mathbf{V})$, represented by a Bayesian network. We assume that from this network a junction tree is constructed, with a set $\mathbf{Cl}$ of cliques and a set $\mathbf{S}$ of separators. The set of variables of the clique $Cl_i \in \mathbf{Cl}$ is denoted as $\mathbf{V}_i = \mathbf{C}_i \cup \mathbf{X}_i$; the variable set of a separator $S_{ij} \in \mathbf{S}$ is indicated by $\mathbf{V}_{ij}$. Each clique Cl_i is supplemented with a marginal distribution $\Pr(\mathbf{V}_i)$; a separator S_{ij} between two cliques Cl_i and Cl_j has associated the distribution $\Pr(\mathbf{V}_{ij}) = \Pr(\mathbf{V}_i \cap \mathbf{V}_j)$. The joint probability distribution $\Pr(\mathbf{V})$ over all variables $\mathbf{V}$ is known to factorise over the junction tree as

$$\Pr(\mathbf{V}) = \frac{\prod_{Cl_i \in \mathbf{Cl}} \Pr(\mathbf{V}_i)}{\prod_{S_{ij} \in \mathbf{S}} \Pr(\mathbf{V}_{ij})}$$

The well-known junction-tree propagation algorithm provides for ef-

[1] Department of Information and Computing Sciences, Utrecht University, The Netherlands; email: {A.J.Pastink, L.C.vanderGaag}@uu.nl

ficiently computing marginal probabilities over the network's variables by means of local message-passing through the junction tree.

The problem in monitoring applications now is to find, at each time t, a joint value combination $\mathbf{c}$ for the network's explanatory variables $\mathbf{C}$ which maximises the posterior probability $\Pr(\mathbf{c} \mid \mathbf{x}^t)$ given the observation vector $\mathbf{x}^t$; such a maximising value combination $\mathbf{c}$ is termed a most probable explanation for $\mathbf{x}^t$. More formally, the most probable explanation for $\mathbf{x}^t$ at time t, denoted by $\mathbf{m}^t$, equals

$$\mathbf{m}^t = \underset{\mathbf{c} \in \Omega_{\mathbf{C}}}{\operatorname{argmax}} \ \Pr(\mathbf{c} \mid \mathbf{x}^t)$$

The explanation $\mathbf{m}^t$ assigns values to the explanatory variables of the separate cliques in the junction tree. For each clique $Cl_i \in \mathbf{Cl}$, the value combination $\mathbf{m}_i^t \in \Omega_{\mathbf{C}_i}$ consistent with $\mathbf{m}^t$ is termed the constituent explanation from Cl_i for the overall explanation $\mathbf{m}^t$.

The problem of finding most probable explanations for observation vectors is known to be NP-hard for Bayesian networks in general [3]. For networks of bounded treewidth, the problem is solvable in polynomial time. To this end, an efficient algorithm called max-propagation is available [1], which builds upon the same concept of local message passing as the standard junction-tree algorithm.

3 LOCAL PROPERTIES OF CONSTITUENTS

We consider a Bayesian network for a monitoring application. Within the network's junction tree, we focus on a single clique Cl_r and study properties of its constituent explanation. We will show more specifically that, while a change of observation elsewhere in the junction tree may induce a change of constituent explanation for Cl_r, the new constituent cannot be any arbitrary value combination from $\Omega_{\mathbf{C}_r}$. Without loss of generality, we assume that Cl_r has the two neighbouring cliques Cl_p and Cl_q, as shown in Figure 1; S_{pr} and S_{rq} are the two clique separators. For ease of exposition, we assume that the separator S_{pr} includes a single explanatory variable P, that is, $S_{pr} = \mathbf{C}_p \cap \mathbf{C}_r = \{P\}$; similarly, we assume that $S_{rq} = \{Q\}$.

We suppose that, given the observation vector $\mathbf{x}^t$ at time t, an overall most probable explanation $\mathbf{m}^t$ for $\mathbf{x}^t$ has been computed from the junction tree. By definition, this explanation maximises the probability $\Pr(\mathbf{c} \mid \mathbf{x}^t)$ over all joint value combinations $\mathbf{c} \in \Omega_{\mathbf{C}}$, that is,

$$\mathbf{m}^t = \underset{\mathbf{c} \in \Omega_{\mathbf{C}}}{\operatorname{argmax}} \left[\frac{\Pr(\mathbf{c}_{p-} \mid \mathbf{x}^t) \cdot \Pr(\mathbf{c}_r \mid \mathbf{x}^t) \cdot \Pr(\mathbf{c}_{q-} \mid \mathbf{x}^t)}{\Pr(p' \mid \mathbf{x}^t) \cdot \Pr(q' \mid \mathbf{x}^t)} \right]$$

where the set $\mathbf{C}_{p-}$ includes the explanatory variables from all cliques that are separated from Cl_r by S_{pr}, and $\mathbf{C}_{q-}$ is defined analogously; the value combinations $\mathbf{c}_{p-}, \mathbf{c}_r, \mathbf{c}_{q-}$ and the values p', q' are taken consistent with $\mathbf{c}$. We note that the value combination $\mathbf{c}_{p-}$ shares its value for the separator variable P with the value combination $\mathbf{c}_r$; it does not share any other values with $\mathbf{c}_r$. A similar property holds for the value combination $\mathbf{c}_{q-}$ and the separator variable Q. We find that the four possible value combinations for the two separator variables P and Q partition the set $\Omega_{\mathbf{C}_r}$ of clique Cl_r into four blocks. The block denoted by $\Omega^{pq}_{\mathbf{C}_r}$ includes all value combinations from $\Omega_{\mathbf{C}_r}$ that are consistent with the value p for P and the value q for Q; the other three blocks $\Omega^{p\bar{q}}_{\mathbf{C}_r}$, $\Omega^{\bar{p}q}_{\mathbf{C}_r}$ and $\Omega^{\bar{p}\bar{q}}_{\mathbf{C}_r}$ have analogous meanings.

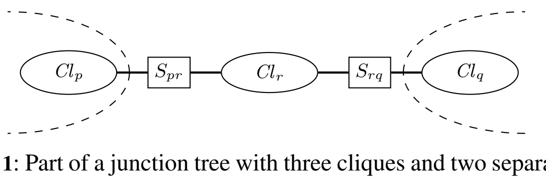


Figure 1: Part of a junction tree with three cliques and two separators.

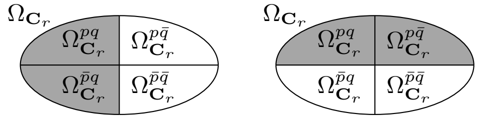

(a) Consistent with q (b) Consistent with p

Figure 2: Partition of the set $\Omega_{\mathbf{C}_r}$ of clique Cl_r into four blocks.

The partition of the set $\Omega_{\mathbf{C}_r}$ of value combinations for the explanatory variables from Cl_r is visualised in Figure 2.

Without further knowledge from the rest of the junction tree, we cannot establish the contribution $\mathbf{m}_r^t$ from clique Cl_r to the overall explanation at time t. We know however that this constituent is one of the most likely value combinations per block, that is,

$$\begin{aligned}\mathbf{m}_r^t \in & \left\{ \underset{\mathbf{c}_r \in \Omega^{pq}_{\mathbf{C}_r}}{\operatorname{argmax}} \Pr(\mathbf{c}_r \mid \mathbf{x}^t),\ \underset{\mathbf{c}_r \in \Omega^{p\bar{q}}_{\mathbf{C}_r}}{\operatorname{argmax}} \Pr(\mathbf{c}_r \mid \mathbf{x}^t), \right. \\ & \left. \underset{\mathbf{c}_r \in \Omega^{\bar{p}q}_{\mathbf{C}_r}}{\operatorname{argmax}} \Pr(\mathbf{c}_r \mid \mathbf{x}^t),\ \underset{\mathbf{c}_r \in \Omega^{\bar{p}\bar{q}}_{\mathbf{C}_r}}{\operatorname{argmax}} \Pr(\mathbf{c}_r \mid \mathbf{x}^t) \right\} \\ & = \{\mathbf{c}_r^{pq}, \mathbf{c}_r^{p\bar{q}}, \mathbf{c}_r^{\bar{p}q}, \mathbf{c}_r^{\bar{p}\bar{q}}\}\end{aligned}$$

where $\mathbf{c}_r^{pq}$ denotes the most likely element from the set $\Omega^{pq}_{\mathbf{C}_r}$ of value combinations with pq, and $\mathbf{c}_r^{\bar{p}q}$, $\mathbf{c}_r^{p\bar{q}}$ and $\mathbf{c}_r^{\bar{p}\bar{q}}$ are defined analogously.

We now suppose that, at time $t+1$, a new value is obtained for some variable in clique Cl_q; all other observable variables retain their original values. As a result of the new observation, the marginal distribution $\Pr(Q \mid \mathbf{x}^t)$ over the separator variable Q is updated to $\Pr(Q \mid \mathbf{x}^{t+1})$. Upon further propagation of the update to clique Cl_r, the probability distribution $\Pr(\mathbf{C}_r \mid \mathbf{x}^t)$ over the explanatory variables from Cl_r is multiplied by update factors such that

$$\Pr(\mathbf{c}_r \mid \mathbf{x}^{t+1}) \quad = \quad \Pr(\mathbf{c}_r \mid \mathbf{x}^t) \cdot \frac{\Pr(q' \mid \mathbf{x}^{t+1})}{\Pr(q' \mid \mathbf{x}^t)}$$

for all value combinations $\mathbf{c}_r \in \Omega_{\mathbf{C}_r}$ and q' consistent with $\mathbf{c}_r$. We now show that the new constituent explanation $\mathbf{m}_r^{t+1}$ from clique Cl_r at time $t+1$ is again among the candidates which were identified for the constituent explanation $\mathbf{m}_r^t$ at time t. We consider to this end the most likely value combination $\mathbf{c}_r^{pq}$ from the block Ω_r^{pq} of clique Cl_r at time t. For this value combination we have that

$$\Pr(\mathbf{c}_r^{pq} \mid \mathbf{x}^t) \quad \geq \quad \Pr(\mathbf{c}'_r \mid \mathbf{x}^t)$$

for all value combinations $\mathbf{c}'_r \in \Omega^{pq}_{\mathbf{C}_r}$ consistent with pq. At time $t+1$, we find for any such combination $\mathbf{c}'_r$ that

$$\begin{aligned}\Pr(\mathbf{c}_r^{pq} \mid \mathbf{x}^{t+1}) \quad &= \quad \Pr(\mathbf{c}_r^{pq} \mid \mathbf{x}^t) \cdot \frac{\Pr(q \mid \mathbf{x}^{t+1})}{\Pr(q \mid \mathbf{x}^t)} \quad \geq \\ \Pr(\mathbf{c}'_r \mid \mathbf{x}^t) \cdot \frac{\Pr(q \mid \mathbf{x}^{t+1})}{\Pr(q \mid \mathbf{x}^t)} \quad &= \quad \Pr(\mathbf{c}'_r \mid \mathbf{x}^{t+1})\end{aligned}$$

We conclude that since all joint value combinations $\mathbf{c}'_r \in \Omega^{pq}_{\mathbf{C}_r}$ are multiplied by the same update factor, $\mathbf{c}_r^{pq}$ remains to be the most likely element from $\Omega^{pq}_{\mathbf{C}_r}$. Similar considerations hold for the most likely value combinations $\mathbf{c}_r^{p\bar{q}}$, $\mathbf{c}_r^{\bar{p}q}$ and $\mathbf{c}_r^{\bar{p}\bar{q}}$ of the other three blocks. Given a single new observation in its neighbouring clique Cl_q, we thus find for clique Cl_r that:

$$\mathbf{m}_r^t \in \{\mathbf{c}_r^{pq}, \mathbf{c}_r^{p\bar{q}}, \mathbf{c}_r^{\bar{p}q}, \mathbf{c}_r^{\bar{p}\bar{q}}\} \ \longrightarrow \ \mathbf{m}_r^{t+1} \in \{\mathbf{c}_r^{pq}, \mathbf{c}_r^{p\bar{q}}, \mathbf{c}_r^{\bar{p}q}, \mathbf{c}_r^{\bar{p}\bar{q}}\}$$

While the new observation vector $\mathbf{x}^{t+1}$ may induce a change in the constituent explanation from Cl_r therefore, its constituent for $\mathbf{m}^{t+1}$ cannot be any arbitrary joint value combination from $\Omega_{\mathbf{C}_r}$.

The result stated above is readily generalised. A new value, at time $t+1$, for a single observable variable in clique Cl_p for example, would lead to a similar result. The result further holds for cliques Cl_r with multiple adjoining separators. The set of value combinations $\Omega_{\mathbf{C}_r}$ for the explanatory variables from such a clique would then be partitioned into as many blocks as there are value combinations for its explanatory separator variables. The result also holds for multiple subsequent changes to the observation vector, as long as these changes do not pertain to the observable variables of the clique Cl_r at hand: since all value combinations from a single block of $\Omega_{\mathbf{C}_r}$ will always be updated by the same factor, regardless of the number of changes to the separators' probability distributions, the most likely combinations per block remain to be the candidates for the constituent explanation from clique Cl_r.

Thus far we studied the effect that a change of value for an observable variable in its neighbouring clique Cl_q can have on the constituent explanation from clique Cl_r. We now briefly address the effect that this change can have on the constituent explanation from clique Cl_p upon propagation further down the junction tree. Referring again to Figure 1, we note that clique Cl_p is linked to the source Cl_q from which the probability update originates, only through Cl_r. We now look upon the two cliques Cl_p and Cl_r as constituting a single joint clique $Cl_{(p,r)}$ with the marginal probability distribution $\Pr(\mathbf{C}_p, \mathbf{C}_r \mid \mathbf{x}^t)$ over its variables established as

$$\Pr(\mathbf{c}_p, \mathbf{c}_r \mid \mathbf{x}^t) = \frac{\Pr(\mathbf{c}_p \mid \mathbf{x}^t) \cdot \Pr(\mathbf{c}_r \mid \mathbf{x}^t)}{\Pr(p' \mid \mathbf{x}^t)}$$

for all values p' of P and all value combinations $\mathbf{c}_p\mathbf{c}_r \in \Omega_{\mathbf{C}_p} \times \Omega_{\mathbf{C}_r}$ consistent with p'. The joint clique $Cl_{(p,r)}$ contributes a constituent explanation $\mathbf{m}^t_{(p,r)} = \mathbf{m}^t_p\, \mathbf{m}^t_r$ composed of the separate constituents from its original cliques, to the overall most probable explanation. This joint constituent is an element of the set $\Omega_{\mathbf{C}_{(p,r)}} = (\Omega^p_{\mathbf{C}_p} \times \Omega^p_{\mathbf{C}_r}) \cup (\Omega^{\bar{p}}_{\mathbf{C}_p} \times \Omega^{\bar{p}}_{\mathbf{C}_r})$ of value combinations for the explanatory variables from both cliques, with $\mathbf{m}^t_p \in \Omega^p_{\mathbf{C}_p} \cup \Omega^{\bar{p}}_{\mathbf{C}_p}$ and $\mathbf{m}^t_r \in \Omega^p_{\mathbf{C}_r} \cup \Omega^{\bar{p}}_{\mathbf{C}_r}$. Assuming, without loss of generality, that $Cl_{(p,r)}$ does not have any neighbouring cliques other than Cl_q, we have that S_{rq} is its only adjacent separator. The separator variable Q thus partitions the set $\Omega_{\mathbf{C}_{(p,r)}}$ into two blocks of value combinations, consistent with q and with $\bar{q}$ respectively. By similar arguments as above, it is now readily seen that a value change for an observable variable from clique Cl_q may induce a change in the constituent explanation from $Cl_{(p,r)}$ and hence in that from Cl_p. If the constituent explanation from Cl_r does not change as a result of the new observation however, then the constituent from Cl_p will not change either.

4 PERSISTENCE OF CONSTITUENTS

In the previous section we focused on a single clique in a junction tree and studied the effect that a new observation elsewhere in the tree can have on its candidate constituents. We argued that although a clique's current constituent may change as a result of the new observation, it cannot change to any arbitrary value combination. While we could establish the candidate constituents for the new overall explanation, we could not decide whether a clique's current constituent would persist, as for establishing persistence information from the rest of the junction tree is required. In this section we investigate properties of persistence of constituents over time. More specifically, we derive conditions under which persistence of a constituent is guaranteed. In Section 5 we will then build upon these conditions to arrive at a tailored algorithm that halts the propagation of probability updates as soon as constituents are known to persist.

4.1 Persistence after a single probability update

We consider the junction tree from Figure 1, with the cliques Cl_p, Cl_r, Cl_q and the separators S_{pr}, S_{rq} as before. Without loss of generality, we assume that clique Cl_p has no neighbouring cliques other than Cl_r. At time t, the most probable explanation $\mathbf{m}^t$ for the observation vector $\mathbf{x}^t$ has been computed; we assume that $\mathbf{m}^t$ includes the value combination pq for the separator variables P and Q.

At time $t+1$, a new value is observed for some variable in clique Cl_q. The most probable explanation for the new observation vector $\mathbf{x}^{t+1}$ is $\mathbf{m}^{t+1}$. Joining the cliques Cl_p and Cl_r into $Cl_{(p,r)}$ as detailed in Section 3, we find for $\mathbf{m}^{t+1}$ that

$$\begin{aligned}\mathbf{m}^{t+1} &= \operatorname*{argmax}_{\mathbf{c}\in\Omega_{\mathbf{C}}} \left[\frac{\Pr(\mathbf{c}_{(p,r)} \mid \mathbf{x}^{t+1})}{\Pr(q' \mid \mathbf{x}^{t+1})} \cdot \Pr(\mathbf{c}_{q-} \mid \mathbf{x}^{t+1}) \right] \\ &= \operatorname*{argmax}_{\mathbf{c}\in\Omega_{\mathbf{C}}} \left[\frac{\Pr(\mathbf{c}_{(p,r)} \mid \mathbf{x}^{t})}{\Pr(q' \mid \mathbf{x}^{t})} \cdot \Pr(\mathbf{c}_{q-} \mid \mathbf{x}^{t+1}) \right]\end{aligned}$$

where the set $\mathbf{C}_{q-}$ is as before, and $\mathbf{c}_{(p,r)}, \mathbf{c}_{q-}, q'$ are taken consistent with $\mathbf{c}$. To study the relation between $\mathbf{m}^{t+1}$ and the most probable explanation $\mathbf{m}^t$ from time t, we distinguish between the two cases $\mathbf{m}^{t+1} \in \Omega^{pq}_{\mathbf{C}}$ and $\mathbf{m}^{t+1} \notin \Omega^{pq}_{\mathbf{C}}$. In the former case, we have that

$$\mathbf{m}^{t+1} = \operatorname*{argmax}_{\mathbf{c}\in\Omega^{pq}_{\mathbf{C}}} \left[\frac{\Pr(\mathbf{c}_{(p,r)} \mid \mathbf{x}^{t})}{\Pr(q \mid \mathbf{x}^{t})} \cdot \Pr(\mathbf{c}_{q-} \mid \mathbf{x}^{t+1}) \right]$$

with $\mathbf{c}_{(p,r)}$, $\mathbf{c}_{q-}, q$ again consistent with $\mathbf{c}$. Since the probabilities of all value combinations $\mathbf{c}_{(p,r)}$ from the set $\Omega^q_{\mathbf{C}_{(p,r)}}$ are multiplied by the same factor $\Pr(\mathbf{c}_{q-} \mid \mathbf{x}^{t+1}) / \Pr(q \mid \mathbf{x}^t)$, we conclude for the constituent explanation $\mathbf{m}^{t+1}_{(p,r)}$ from $Cl_{(p,r)}$ at time $t+1$ that

$$\mathbf{m}^{t+1}_{(p,r)} = \operatorname*{argmax}_{\mathbf{c}\in\Omega^{pq}_{\mathbf{C}_{(p,r)}}} \Pr(\mathbf{c}_{(p,r)} \mid \mathbf{x}^t) = \mathbf{m}^t_{(p,r)}$$

that is, we find that the constituent explanation from time t persists onto time $t+1$; the separate constituents from the original cliques Cl_p and Cl_r thus persist as well. In the case where $\mathbf{m}^{t+1} \notin \Omega^{pq}_{\mathbf{C}}$, we know that the constituent explanation $\mathbf{m}^t_{(p,r)}$ from the joint clique $Cl_{(p,r)}$ at time t does not persist onto time $t+1$. In fact, we know that at least the separator variable Q will have changed value in the new constituent $\mathbf{m}^{t+1}_{(p,r)}$. We derive that

$$\mathbf{m}^{t+1}_{(p,r)} = \operatorname*{argmax}_{\mathbf{c}\in\Omega^{\bar{q}}_{\mathbf{C}_{(p,r)}}} \Pr(\mathbf{c}_{(p,r)} \mid \mathbf{x}^t)$$

that is, the new constituent explanation from the joint clique $Cl_{(p,r)}$ is the most likely value combination given $\mathbf{x}^t$ from the block $\Omega^{\bar{q}}_{\mathbf{C}_{(p,r)}}$.

From the considerations above, we have that the overall explanation $\mathbf{m}^t$ with the value combination pq for the separator variables P and Q, persists onto time $t+1$ if the following inequality holds

$$\max_{\mathbf{c}\in\Omega^{pq}_{\mathbf{C}}} \left[\Pr(\mathbf{c}_{(p,r)} \mid \mathbf{x}^t) \cdot \frac{\Pr(\mathbf{c}_{q-} \mid \mathbf{x}^{t+1})}{\Pr(q \mid \mathbf{x}^t)} \right] \geq \max_{\mathbf{c}'\in\Omega^{\bar{q}}_{\mathbf{C}}} \left[\Pr(\mathbf{c}'_{(p,r)} \mid \mathbf{x}^t) \cdot \frac{\Pr(\mathbf{c}'_{q-} \mid \mathbf{x}^{t+1})}{\Pr(\bar{q} \mid \mathbf{x}^t)} \right]$$

where $\mathbf{c}_{(p,r)}, \mathbf{c}_{q-}, q$ are consistent with $\mathbf{c}$, and $\mathbf{c}'_{(p,r)}, \mathbf{c}'_{q-}, \bar{q}$ are consistent with $\mathbf{c}'$; in the sequel, we will use the phrase persistence inequality to refer to this inequality. Since

$$\operatorname*{argmax}_{\mathbf{c}\in\Omega^{pq}_{\mathbf{C}}} \left[\Pr(\mathbf{c}_{(p,r)} \mid \mathbf{x}^t) \cdot \frac{\Pr(\mathbf{c}_{q-} \mid \mathbf{x}^{t+1})}{\Pr(q \mid \mathbf{x}^t)} \right] = \mathbf{m}^t$$

we know that the joint value combination $\mathbf{c}_{(p,r)}$ maximizing the left-hand side of the persistence inequality equals the current constituent explanation $\mathbf{m}^t_{(p,r)}$ from the joint clique $Cl_{(p,r)}$. The right-hand side of the inequality is maximised by one of the candidate constituents $\mathbf{c}^{p\bar{q}}_{(p,r)}, \mathbf{c}^{\bar{p}\bar{q}}_{(p,r)}$, identified for clique $Cl_{(p,r)}$ at time t.

To summarise, we found that, given a value change in its neighbouring clique Cl_q, the constituent explanation $\mathbf{m}^t_r$ from clique Cl_r at time t is guaranteed to persist onto time $t+1$ if and only if the persistence inequality holds. In Section 5 we will argue that the probabilistic information required for verifying the inequality is available locally in the separators of the junction tree upon runtime, and hence can be used to decide whether or not propagation can be halted.

4.2 Persistence after multiple probability updates

Having gained insight in the persistence of constituent explanations after a single probability update, we now address persistence after multiple consecutive updates. For this purpose, we distinguish between two types of persistence for separator constituents. For two cliques Cl_i and Cl_j separated by the explanatory variable K, we say that the value k for K at time t persists strongly after a probability update in clique Cl_j, if k persists in the overall explanation to time $t+1$, and in addition the following inequality holds:

$$\frac{\max_{\mathbf{c}_{j-}\in\Omega^k_{\mathbf{C}_{j-}}} \Pr(\mathbf{c}_{j-}|\mathbf{x}^{t+1})}{\max_{\mathbf{c}_{j-}\in\Omega^k_{\mathbf{C}_{j-}}} \Pr(\mathbf{c}_{j-}|\mathbf{x}^{t})} \geq \frac{\max_{\mathbf{c}'_{j-}\in\Omega^{\bar{k}}_{\mathbf{C}_{j-}}} \Pr(\mathbf{c}'_{j-}|\mathbf{x}^{t+1})}{\max_{\mathbf{c}'_{j-}\in\Omega^{\bar{k}}_{\mathbf{C}_{j-}}} \Pr(\mathbf{c}'_{j-}|\mathbf{x}^{t})}$$

where $\mathbf{C}_{j-}$ is the set of explanatory variables separated from Cl_i by K; $\mathbf{c}_{j-}$ is taken consistent with k, and $\mathbf{c}'_{j-}$ includes $\bar{k}$. If the separator value k persists from time t to time $t+1$ yet not strongly so, we say that its persistence is weak. We will show that if at most one of the separator constituents for a clique persists weakly after an associated value change, then the clique's constituent explanation will persist after multiple value changes throughout the junction tree.

We consider as before the junction tree from Figure 1, with the cliques Cl_p, Cl_r, Cl_q and the separators S_{pr}, S_{rq}. At time t, the value combination $\mathbf{m}^t$ constitutes the most probable explanation for the observation vector $\mathbf{x}^t$; we assume again that $\mathbf{m}^t$ includes the value combination pq for the two separator variables, and hence that $\mathbf{m}^t \in \Omega^{pq}_{\mathbf{C}}$. We now consider two separate changes to the observation vector $\mathbf{x}^t$. One of these changes pertains to an observable variable from clique Cl_q; we write $\mathbf{x}^{(Q)}$ to denote the resulting observation vector. The other value change takes place in clique Cl_p; we write $\mathbf{x}^{(P)}$ to denote the observation vector resulting from just this change. We now suppose that the two changes to the observation vector $\mathbf{x}^t$ are effectuated consecutively. The first change gives the observation vector $\mathbf{x}^{t+1} = \mathbf{x}^{(Q)}$, and the vector after both changes is $\mathbf{x}^{t+2}$; note that $\mathbf{x}^{t+2} = \mathbf{x}^{(P\circ Q)} \neq \mathbf{x}^{(P)}$. We would like to mention that the described situation can arise upon runtime with a propagation algorithm that after a single value change verifies persistence locally and halts as soon as separator persistence is guaranteed. We will show for such a situation that, given persistence of q and strong persistence of p in view of the original observation vector $\mathbf{x}^t$, the probability of the most likely value combination from $\Omega^{pq}_{\mathbf{C}}$ at time $t+2$ is larger than that from $\Omega^{\bar{p}\bar{q}}_{\mathbf{C}}$. Since similar properties also hold for the most likely value combinations given $\mathbf{x}^{t+2}$ from the blocks $\Omega^{p\bar{q}}_{\mathbf{C}}$ and $Omega^{\bar{p}q}_{\mathbf{C}}$, we can conclude persistence of $\mathbf{m}^t_r$ from clique Cl_r onto time $t+2$.

We suppose that the separator value q is known to persist after the first value change in view of the original observation vector $\mathbf{x}^t$. From our considerations in Section 4.1, we have that the most probable explanation $\mathbf{m}^{(Q)}$ for the new observation vector $\mathbf{x}^{(Q)}$ again is an element of the block $\Omega^{pq}_{\mathbf{C}}$ of value combinations including pq. For the new overall explanation $\mathbf{m}^{(Q)}$, we have that

$$\Pr(\mathbf{m}^{(Q)}|\mathbf{x}^{(Q)}) = \max_{\mathbf{c}\in\Omega^{pq}_{\mathbf{C}}}\left[\frac{\Pr(\mathbf{c}_{p-}|\mathbf{x}^t)\cdot\Pr(\mathbf{c}_r|\mathbf{x}^t)\cdot\Pr(\mathbf{c}_{q-}|\mathbf{x}^{(Q)})}{\Pr(p \mid \mathbf{x}^t)\cdot\Pr(q \mid \mathbf{x}^t)}\right] \geq \max_{\mathbf{c}'\in\Omega^{\bar{p}\bar{q}}_{\mathbf{C}}} \Pr(\mathbf{c}' \mid \mathbf{x}^{(Q)})$$

for all value combinations $\mathbf{c}' \in \Omega^{\bar{p}\bar{q}}_{\mathbf{C}}$, where $\mathbf{c}_{p-}, \mathbf{c}_r, \mathbf{c}_{q-}$ are consistent with $\mathbf{c}$; similar properties hold with respect to the blocks $\Omega^{p\bar{q}}_{\mathbf{C}}, \Omega^{\bar{p}q}_{\mathbf{C}}$. We now consider the second value change, pertaining to clique Cl_p. We suppose that after the probability update in Cl_p, the separator value p persists in view of the original observation vector $\mathbf{x}^t$, and that in fact the inequality for strong persistence holds, that is,

$$\frac{\max_{\mathbf{c}_{p-}\in\Omega^p_{\mathbf{C}_{p-}}} \Pr(\mathbf{c}_{p-}|\mathbf{x}^{(P)})}{\max_{\mathbf{c}_{p-}\in\Omega^p_{\mathbf{C}_{p-}}} \Pr(\mathbf{c}_{p-}|\mathbf{x}^{t})} \geq \frac{\max_{\mathbf{c}'_{p-}\in\Omega^{\bar{p}}_{\mathbf{C}_{p-}}} \Pr(\mathbf{c}'_{p-}|\mathbf{x}^{(P)})}{\max_{\mathbf{c}'_{p-}\in\Omega^{\bar{p}}_{\mathbf{C}_{p-}}} \Pr(\mathbf{c}'_{p-}|\mathbf{x}^{t})}$$

We note that from the two persistence properties, we have that the constituent explanation $\mathbf{m}^t_r$ from clique Cl_r is guaranteed to persist with $\mathbf{x}^{(Q)}$ and with $\mathbf{x}^{(P)}$. By incorporating the property of strong persistence in the expression for the probability $\Pr(\mathbf{m}^{(Q)} \mid \mathbf{x}^{(Q)})$ above, we find that it also persists with $\mathbf{x}^{(P\circ Q)}$ after the two changes:

$$\max_{\mathbf{c}\in\Omega^{pq}_{\mathbf{C}}}\left[\Pr(\mathbf{c}_{p-}|\mathbf{x}^{(P)})\cdot\frac{\Pr(\mathbf{c}_r \mid \mathbf{x}^t)}{\Pr(p \mid \mathbf{x}^t)\cdot\Pr(q \mid \mathbf{x}^t)}\cdot\Pr(\mathbf{c}_{q-}|\mathbf{x}^{(Q)})\right] \geq \max_{\mathbf{c}'\in\Omega^{\bar{p}\bar{q}}_{\mathbf{C}}}\left[\Pr(\mathbf{c}'_{p-}|\mathbf{x}^{(P)})\cdot\frac{\Pr(\mathbf{c}'_r \mid \mathbf{x}^t)}{\Pr(\bar{p} \mid \mathbf{x}^t)\cdot\Pr(\bar{q} \mid \mathbf{x}^t)}\cdot\Pr(\mathbf{c}'_{q-}|\mathbf{x}^{(Q)})\right]$$

where $\mathbf{c}_{p-}, \mathbf{c}_r, \mathbf{c}_{q-}$ are consistent with $\mathbf{c} \in \Omega^{pq}_{\mathbf{C}}$ and $\mathbf{c}'_{p-}, \mathbf{c}'_r, \mathbf{c}'_{q-}$ are consistent with $\mathbf{c}' \in \Omega^{\bar{p}\bar{q}}_{\mathbf{C}}$; similar properties are again found for the blocks $\Omega^{p\bar{q}}_{\mathbf{C}}$ and $\Omega^{\bar{p}q}_{\mathbf{C}}$. We thus have that, given weak persistence of the separator value q and strong persistence of p in view of the original observation vector $\mathbf{x}^t$, the most likely value combination from the block Ω^{pq} remains the largest among the most likely value combinations of all four blocks of $\Omega_{\mathbf{C}}$.

While stated for a clique Cl_r with two adjoining separators, the result is readily generalised to cliques with an arbitrary number of separators, as was also argued in Section 3. The result is further generalised to more than two consecutive updates. For our overall result, we then have that the constituent explanation from a clique persists as long as all update factors applied to the clique's marginal distribution originate from separators of which the value persists and strongly so from all but possibly one separator. The order in which the various update factors are applied to a clique's marginal distribution is irrelevant; also the separator from which a weakly persisting value originates is immaterial. More formal proofs of our statements will be provided in a forthcoming technical paper.

5 EXPLOITING PERSISTENCE PROPERTIES

Having identified conditions under which constituent explanations are guaranteed to persist over time, we now present our propagation algorithm tailored to monitoring applications, which exploits these conditions for halting the propagation of probability updates as soon as constituents are known to persist. By building upon the existing cautious and max-propagation algorithms, our algorithm effectively minimizes the number of cliques visited upon propagation.

5.1 Cautious max-propagation

The junction-tree propagation algorithm for Bayesian networks in general is ill suited for monitoring applications with consecutively changing observations. When an observation x is entered into a clique, the algorithm effectively sets the probabilities of all value combination inconsistent with x to zero, and thereby prohibits the retrieval of the original probabilistic information. Before the new observation $\bar{x}$ can be propagated therefore, the junction tree needs to be re-initialized. For studying the effects of alternative observations, cautious propagation has been proposed as a variant of the junction-tree algorithm which retains the original distributions per clique for future computations [2]. For ease of exposition, we assume that, as with cautious propagation, probabilities given alternative observations are readily accessible for our algorithm.

Also for computing a most probable explanation for a given observation vector has a variant of the standard junction-tree propagation algorithm been designed, called max-propagation [1]. While the standard junction-tree propagation algorithm enforces consistency of taking sums over probabilities to ensure correct marginal distributions per clique, the max-propagation algorithm enforces consistency of taking the maximum of probabilities. More specifically, max-propagation maintains for each value combination $\mathbf{c}_i$ for a clique Cl_i the probability that is maximally attained by a most probable explanation consistent with $\mathbf{c}_i$ and the current observation vector $\mathbf{x}^t$, that is, it maintains the max-distribution $\max_{\mathbf{c}_- \in \Omega_{\mathbf{C}_-}} \Pr(\mathbf{c}_i, \mathbf{c}_-, \mathbf{x}^t)$ for all value combinations $\mathbf{c}_i \in \Omega_{\mathbf{C}_i}$, where $\mathbf{C}_- = \mathbf{C} \setminus \mathbf{C}_i$ is the set of explanatory variables not included in Cl_i. Similar information is maintained per separator. From every clique and each separator therefore, its constituent explanation is readily found by choosing a value combination with maximum probability. Retrieving the overall most probable explanation requires some simple extra bookkeeping.

5.2 The propagation algorithm

Building upon concepts from cautious max-propagation, we designed a new propagation algorithm for computing most probable explanations, tailored to monitoring applications. Upon detailing our algorithm, we will again refer to the junction tree from Figure 1. We assume that the tree has been initialized at time t with the observation vector $\mathbf{x}^t$ and that the most probable explanation $\mathbf{m}^t$ for $\mathbf{x}^t$ includes the value combination pq for the separator variables P and Q. For ease of reference, Algorithm 1 summarises the structure of our algorithm; in Section 5.3 we will illustrate the working of our algorithm.

We suppose that a new value is observed for some variable in clique Cl_q, which results in the new observation vector $\mathbf{x}^{(Q)}$. The max-distribution over clique Cl_q is updated with the new information. From the updated max-distribution, the update factors to be sent to the separator $S_{rq} = \{Q\}$ are established as

$$\left\{ \max_{\mathbf{c} \in \Omega_{\mathbf{C}}^q} \Pr(\mathbf{c}, \mathbf{x}^{(Q)}),\ \max_{\mathbf{c}' \in \Omega_{\mathbf{C}}^{\bar{q}}} \Pr(\mathbf{c}', \mathbf{x}^{(Q)}) \right\}$$

We observe that these factors in essence suffice for verifying the persistence inequality for the separator variable Q. If the inequality holds, propagation is halted. Some cliques in the junction tree will then have incorporated the new observation vector $\mathbf{x}^{(Q)}$, while other cliques still have a max-distribution given a previous observation vector. Computation of the most probable explanation for $\mathbf{x}^{(Q)}$ is nonetheless guaranteed to yield the correct value combination.

We now suppose that a new value is observed for a variable in a clique Cl_p which still has a max-distribution given $\mathbf{x}^t$, that is, upon propagating the observation vector $\mathbf{x}^{(Q)}$ the algorithm identified persistence of constituents before the propagation had reached Cl_p. We use, once more, $\mathbf{x}^{(P)}$ to denote the observation vector which results from incorporating the value change in clique Cl_p in $\mathbf{x}^t$. The algorithm updates the max-distribution from clique Cl_p with the new information and starts an outward max-propagation from the clique. For establishing persistence of the constituent explanation from the clique Cl_r at which both updates convene, we need to verify strong persistence of either of its separator constituents. Since the max-distribution maintained for a separator variable pertains to full value combinations, for verifying the inequality for strong persistence upon runtime, this inequality has to be extended to include value combinations from $\Omega_{\mathbf{C}}$. Given persistence of the separator value p, we have that the probability of the most probable explanation equals

$$\Pr(\mathbf{m}^{(P)} \mid \mathbf{x}^{(P)}) = \max_{\mathbf{c} \in \Omega_{\mathbf{C}}^p} \left[\Pr(\mathbf{c}_{p-} \mid \mathbf{x}^{(P)}) \cdot \frac{\Pr(\mathbf{c}_- \mid \mathbf{x}^t)}{\Pr(p \mid \mathbf{x}^t)} \right]$$

with $\mathbf{C}_- = \mathbf{C} \setminus \mathbf{C}_{p-}$. Since the term $\Pr(\mathbf{c}_- \mid \mathbf{x}^t) / \Pr(p \mid \mathbf{x}^t)$ is the same for all value assignments $\mathbf{c}_{p-}$ given $\mathbf{x}^t$ and $\mathbf{x}^{(P)}$), we find that

$$\frac{\max_{\mathbf{c}_{p-} \in \Omega_{\mathbf{C}_{p-}}^p} \Pr(\mathbf{c}_{p-} \mid \mathbf{x}^{(P)})}{\max_{\mathbf{c}_{p-} \in \Omega_{\mathbf{C}_{p-}}^p} \Pr(\mathbf{c}_{p-} \mid \mathbf{x}^t)} = \frac{\max_{\mathbf{c} \in \Omega_{\mathbf{C}}^p} \Pr(\mathbf{c} \mid \mathbf{x}^{(P)})}{\max_{\mathbf{c} \in \Omega_{\mathbf{C}}^p} \Pr(\mathbf{c} \mid \mathbf{x}^t)}$$

for the left-hand side of the strong-persistence inequality; a similar result holds for its right-hand side. We observe that the factors now required for verifying the inequality are directly available upon max-propagation as the update factors at each separator.

We recall from Section 4.2, that in view of a single separator's weak persistence, local verification of simple inequalities suffices to guarantee persistence of constituent explanations upon halting further propagation. When more than one separator value persists weakly, local verification no longer suffices for this purpose. To identify multiple weak separator persistences, the algorithm broadcasts a message signalling weak persistence throughout the junction tree as soon as a first occurrence of weak persistence is found. Upon finding a second weak persistence, a full propagation is started.

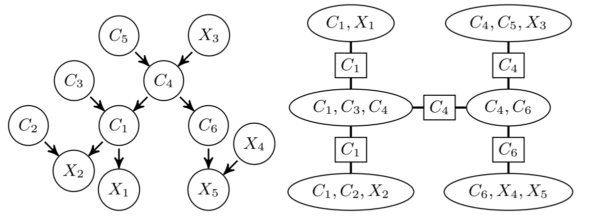
Given a new value for some variable in clique Cl_q, update Cl_q's max-distribution, and start an outward max-propagation;

for *every separator S adjacent to Cl_q* **do**
  if *a marked separator S' exists with $S \perp\!\!\!\perp S' | Cl_q$* **then**
    Propagate from S' to Cl_q and unmark S';
  if *persistence is guaranteed* **then**
    if *the persistence is strong* **then**
      Halt the propagation, and mark S as persisted;
    if *the persistence is the first weak one* **then**
      Halt the propagation, and mark S as persisted;
      Broadcast weak persistence;
    if *the persistence is the second weak one* **then**
      Reset all marks, and start a full propagation;
  else
    if *S is not yet marked as persisted on the other side* **then**
      Continue the propagation;
    else
      Reset all marks, and start a full propagation;
end

Algorithm 1: Summary of propagation for monitoring applications.

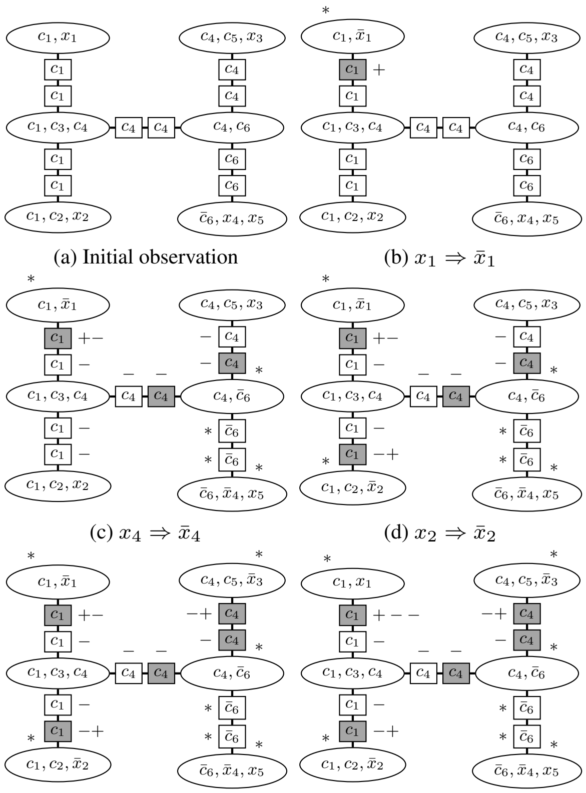
5.3 An illustration of our algorithm

We illustrate the basic idea of our new propagation algorithm for monitoring applications, by means of the example Bayesian network and associated junction tree from Figure 3. We assume that the junction tree is initialized with the observation vector $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$. We further assume that the most probable explanation given $\mathbf{x}$ is established to be $\mathbf{m} = (c_1, c_2, c_3, c_4, c_5, \bar{c}_6)$. Figure 4 now visualizes the run described below. The values mentioned for the variables per clique correspond with the current observation vector and with the most probable explanation given that vector. An '$*$' over a clique or separator indicates that its max-distribution has been updated given a new observation. A '$+$' for a separator indicates that its value has strongly persisted; a '$-$' marks the occurrence of a weak persistence somewhere in the junction tree. Persistence at a separator is indicated by shading. To simplify bookkeeping we split each separator of the junction tree into to parts as shown in Figure 4. Verification of persistence for a separator is performed at the first part that is visited upon propagation.



(a) The example Bayesian network (b) Its associated junction tree

Figure 3: The example network with its corresponding junction tree.



(a) Initial observation (b) $x_1 \Rightarrow \bar{x}_1$

(c) $x_4 \Rightarrow \bar{x}_4$ (d) $x_2 \Rightarrow \bar{x}_2$

(e) $x_3 \Rightarrow \bar{x}_3$ (f) $\bar{x}_1 \Rightarrow x_1$

Figure 4: A run of the algorithm with multiple updates.

We now suppose that the following changes are seen in the consecutive observation vectors entered into the junction tree:

- x_1 is changed to $\bar{x}_1$ — The value change occurs in clique $\{C_1, X_1\}$. We assume that the value c_1 of its separator with $\{C_1, C_3, C_4\}$ persists strongly. The algorithm halts the propagation and marks the separator as having strongly persisted.
- x_4 is changed to $\bar{x}_4$ — The change is entered into $\{C_6, X_4, X_5\}$. We assume that the algorithm identifies, upon propagation, weak persistence of the value c_4 for two separators adjacent to the clique $\{C_4, C_6\}$. Further propagation is halted and the algorithm broadcasts the weak persistence throughout the rest of the junction tree.
- x_2 is changed to $\bar{x}_2$ — We assume strong persistence at c_1, which causes the algorithm to halt further propagation.
- x_3 is changed to $\bar{x}_3$, with strong persistence at c_4.
- $\bar{x}_1$ is changed back again to x_1 – We assume that the change incurs weak persistence of the value c_1 for the separator adjacent to clique $\{C_1, X_1\}$. The second mark of weak persistence now causes the algorithm to invoke a full propagation before any new information is entered.

6 CONCLUSIONS

When employing a Bayesian network for a monitoring application, most probable explanations have to be established for a sequence of consecutive observation vectors. Since computing a single most probable explanation already has an unfavourable runtime complexity, monitoring applications will rapidly become impracticable. In this paper, we studied the computation of a sequence of explanations for subsequent observation vectors which differ by their values for a single variable only. We showed that although a new observation may locally change the explanation, it cannot change to any arbitrary value combination. We have also shown that propagation may be halted as soon as persistence of parts of the explanation are guaranteed. We used these results in a new propagation algorithm tailored to monitoring applications. Our algorithm forestalls unnecessary recomputations of explanations to a large extent, but may in specific situations perform a full propagation through the junction tree. We think it worthwhile to investigate the possibility of performing partial propagations in these situations. Most of all, we plan to study the runtime performance of our algorithm on real-world monitored data.

Acknowledgement. We would like to thank the Netherlands Organisation for Scientific Research for their financial support of the *BioBayes* project.

REFERENCES

[1] A Philip Dawid, 'Applications of a general propagation algorithm for probabilistic expert systems', *Statistics and Computing*, **2**(1), 25–36, (1992).

[2] Finn V Jensen, 'Cautious propagation in Bayesian networks', in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 323–328. Morgan Kaufmann Publishers Inc., (1995).

[3] Solomon Eyal Shimony, 'Finding MAPs for belief networks is NP-hard', *Artificial Intelligence*, **68**, 399 – 410, (1994).