Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

An Exact Algorithm for Solving Most Relevant Explanation in Bayesian Networks

Xiaoyuan Zhu and **Changhe Yuan**
Queens College, City University of New York
65-30 Kissena Blvd., Queens, NY 11367
{xiaoyuan.zhu, changhe.yuan}@qc.cuny.edu

Abstract

Most Relevant Explanation (MRE) is a new inference task in Bayesian networks that finds the most relevant partial instantiation of target variables as an explanation for given evidence by maximizing the Generalized Bayes Factor (GBF). No exact algorithm has been developed for solving MRE previously. This paper fills the void and introduces a breadth-first branch-and-bound MRE algorithm based on a novel upper bound on GBF. The bound is calculated by decomposing the computation of the score to a set of Markov blankets of subsets of evidence variables. Our empirical evaluations show that the proposed algorithm scales up exact MRE inference significantly.

Introduction

Bayesian networks are probabilistic models that represent the conditional independencies between random variables as directed acyclic graphs, and provide principled approaches to explaining evidence, e.g., Most Probable Explanation (MPE), Maximum a Posterior (MAP), and Most Relevant Explanation (MRE). Both MPE and MAP are solved by optimizing the joint posterior probability of the target variables given the evidence. However, in explanation problems (Lacave and Díez 2002; Nielsen, Pellet, and Elisseeff 2008), MPE and MAP often produce overspecified explanations (Pacer et al. 2013; Yuan et al. 2009), i.e., irrelevant target variables may be included. This is because maximizing the joint probability alone cannot exclude irrelevant target variables directly. To address the limitation of MPE and MAP, MRE (Yuan, Lim, and Lu 2011; Yuan and Lu 2008) was proposed to find a partial instantiation of the target variables that maximizes the Generalized Bayes Factor (GBF) (Fitelson 2007; Good 1985) as an explanation for the evidence. GBF is a rational function of probabilities that is suitable for comparing explanations with different cardinalities. Theoretically, MRE is shown able to prune away independent and less relevant variables from the final explanation (Yuan, Lim, and Lu 2011); a recent human study also provided empirical evidence to support the claim (Pacer et al. 2013).

Due to the difficulty of finding a meaningful bound for GBF, no exact algorithms have been developed to solve MRE. Only local search and Markov chain Monte Carlo methods have been proposed previously (Yuan, Lim, and Littman 2011). In this paper, we introduce the first non-trivial exact MRE algorithm, a Breadth-First Branch-and-Bound (BFBnB) algorithm. The key idea of the proposed method is to decompose the whole Bayesian network into a set of overlapping subnetworks. Each subnetwork is characterized by a subset of evidence variables and its Markov blanket. An upper bound for the GBF of an MRE solution is derived by solving independent optimization problems in the subnetworks. To tighten the upper bound, we also introduce a greedy algorithm for merging the Markov blankets that share target variables. The upper bound is used in the BFBnB algorithm for pruning the search space. We evaluate the proposed algorithm in a set of benchmark diagnostic Bayesian networks. The experimental results show that the BFBnB algorithm scales up exact MRE inference significantly.

Background

A Bayesian network (Pearl 1988; Koller and Friedman 2009) is represented as a Directed Acyclic Graph (DAG). The nodes in the DAG represent random variables. The lack of arcs in the DAG define conditional independence relations among the nodes. If there is an arc from node Y to X, we say that Y is a parent of X, and X is a child of Y (We use upper-case letters to denote variables X or variable sets $\mathbf{X}$, and lower-case letters for values of scalars x or vectors $\mathbf{x}$.). A node Y is an ancestor of a node X, if there is a directed path from Y to X. Let $an(X)$ denote all the ancestors of X, then the smallest ancestral set $An(\mathbf{X})$ of node set $\mathbf{X}$ is defined as $An(\mathbf{X}) = \mathbf{X} \cup (\cup_{X_i \in \mathbf{X}} an(X_i))$. *D-separation* describes the conditional independence relation between two set of nodes $\mathbf{X}$ and $\mathbf{Y}$, given a third set of nodes $\mathbf{Z}$, i.e., $p(\mathbf{X}|\mathbf{Z}, \mathbf{Y}) = p(\mathbf{X}|\mathbf{Z})$. The Markov blanket of $\mathbf{X}$, denoted by $MB(\mathbf{X})$, is the smallest node set which d-separates $\mathbf{X}$ from the remaining nodes in the network. The network as a whole represents the joint probability distribution of $\prod_X p(X|PA(X))$, where $PA(X)$ is the set of all the parents of X.

The moral graph G^m of a DAG G is an undirected graph with the same set of nodes. There is an edge between X and

Y in G^m if and only if there is an edge between them in G or if they are parents of the same node in G. In an undirected graph, $\mathbf{Z}$ separates $\mathbf{X}$ and $\mathbf{Y}$, if $\mathbf{Z}$ intercepts all paths between $\mathbf{X}$ and $\mathbf{Y}$. Moral graph is a powerful construction to explain d-separation. Lemma 1 (Lauritzen et al. 1990) links d-separation in DAG to separation in undirected graphs.

Lemma 1. *Let $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ be disjoint subsets of nodes in a DAG G. Then $\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$ if and only if $\boldsymbol{Z}$ separates $\boldsymbol{X}$ from $\boldsymbol{Y}$ in $(G_{An(\boldsymbol{X}\cup\boldsymbol{Y}\cup\boldsymbol{Z})})^m$, where $(G_{An(\boldsymbol{X}\cup\boldsymbol{Y}\cup\boldsymbol{Z})})^m$ is the moral graph of the subgraph of G with node set $An(\boldsymbol{X}\cup\boldsymbol{Y}\cup\boldsymbol{Z})$.*

In an inference problem in a Bayesian network, the nodes are often classified into three categories: target, evidence, and auxiliary. The target set $\mathbf{M}$ represents variables of inference interest. The evidence set $\mathbf{E}$ represents observed information. The auxiliary set represents the variables that are not of interest in the inference. MRE (Yuan, Lim, and Lu 2011) finds a partial instantiation of $\mathbf{M}$ as an explanation for given evidence $\mathbf{e}$ in a Bayesian network. Here, *explanation* refers to the explanation of evidence, whose goal is to explain why some observed variables are in their particular states using the target variables in the domain. The search space of MRE is large and the complexity of MRE is conjectured to be NP^{PP} complete (Yuan, Lim, and Littman 2011). Formally, the MRE problem is formulated as follows:

Definition 1. *Let $\mathbf{M}$ be a set of targets, and $\mathbf{e}$ be the given evidence in a Bayesian network. Most Relevant Explanation is the problem of finding a partial instantiation $\mathbf{x}$ of $\mathbf{M}$ that has the maximum generalized Bayes factor score $GBF(\mathbf{x};\mathbf{e})$ as the explanation for $\mathbf{e}$, i.e.,*

$$MRE(\mathbf{M};\mathbf{e}) = \operatorname*{argmax}_{\mathbf{x},\emptyset\subset\mathbf{X}\subseteq\mathbf{M}} GBF(\mathbf{x};\mathbf{e}), \tag{1}$$

where GBF is defined as

$$GBF(\mathbf{x};\mathbf{e}) = \frac{p(\mathbf{e}|\mathbf{x})}{p(\mathbf{e}|\bar{\mathbf{x}})}. \tag{2}$$

Here, $\mathbf{x}$ is the instantiation of $\mathbf{X}$, and $\bar{\mathbf{x}}$ represents all of the alternative explanations of $\mathbf{x}$.

Different from MPE and MAP which maximize a joint posterior probability function, MRE maximizes the rational function of probabilities in Equation 2. This makes it possible for MRE to compare explanations with different cardinalities in a principled way.

Belief update ratio is a useful concept. The belief update ratio of $\mathbf{X}$ given $\mathbf{e}$ is defined as follows:

$$r(\mathbf{X};\mathbf{e}) = \frac{p(\mathbf{X}|\mathbf{e})}{p(\mathbf{X})}. \tag{3}$$

GBF can be calculated from the belief update ratio as follows:

$$\begin{aligned} GBF(\mathbf{x};\mathbf{e}) &= \frac{p(\mathbf{x}|\mathbf{e})(1-p(\mathbf{x}))}{p(\mathbf{x})(1-p(\mathbf{x}|\mathbf{e}))} = \frac{r(\mathbf{x};\mathbf{e})-p(\mathbf{x}|\mathbf{e})}{1-p(\mathbf{x}|\mathbf{e})} \\ &= 1+\frac{r(\mathbf{x};\mathbf{e})-1}{1-p(\mathbf{x}|\mathbf{e})}. \end{aligned} \tag{4}$$

Many local search methods have been applied to solve both MAP/MPE (Park and Darwiche 2001) and MRE (Yuan, Lim, and Littman 2011), such as tabu search (Glover 1990). Tabu search starts at an empty solution set. At each step, it generates the neighbors of the current solution by adding, changing, or deleting one target variable. Then tabu search selects the best neighbor which has the highest GBF score and has not been visited before. In tabu search, the best neighbor can be worse than the current solution. To stop tabu search properly, upper bounds are set on both the total number of search steps M and the number of search steps since the last improvement L as the stopping criteria. Local search methods can only provide approximate solutions which are not guaranteed to be optimal. Furthermore, the accuracy and efficiency of these methods are typically highly sensitive to tunable parameters.

Branch-and-bound algorithms have been developed for solving MAP and MPE by using upper bounds derived based on different relaxations. A mini-bucket upper bound is proposed in (Dechter and Rish 2003) and applied to And/Or tree search for solving MPE (Marinescu and Dechter 2009). The work in (Choi, Chavira, and Darwiche 2007) showed that the mini-bucket upper bound can be derived from a node splitting scheme. To solve MAP exactly, an upper bound is proposed in (Park and Darwiche 2003) by commuting the order of max and sum operations in the MAP calculation. In (Huang, Chavira, and Darwiche 2006), an exact algorithm is proposed for solving MAP by computing upper bounds on an arithmetic circuit compiled from a Bayesian network. No upper bound or exact algorithm has been proposed for solving MRE, however.

An Exact MRE Algorithm Based on A Novel Upper Bound

It is difficult to solve MRE problems exactly because of both an exponential search space and the need for probabilistic inference at each search step. A naive brute-force search method can scale to Bayesian networks with at most 15 targets. In this work, we propose a breadth-first branch-and-bound algorithm that uses a new upper bound based on Markov blanket decomposition to prune the search space. The algorithm makes it possible to solve MRE problems with more targets exactly.

Search space formulation

Assuming there are n targets, and each target has d states, the search space of MRE contains $(d+1)^n-1$ possible joint states (or solutions). Each state contains values of a subset of the targets. We organize the search space as a search tree by instantiating the targets according to a total order π of the targets. The search tree has the empty state as the root. For a state $\mathbf{y}$ in the tree, $\mathbf{Y}$ is defined as the *expanded* set of targets, and $\mathbf{X} = \{X|X \in \mathbf{M};\ \forall Y_i \in \mathbf{Y},\ Y_i <_\pi X\}$ is the *unexpanded* set. That is, the unexpanded set only contains targets that trail all of the expanded targets in π. Each child state of $\mathbf{y}$ instantiates one more unexpanded target. Figure 1 shows an example with two targets $A = \{a, \bar{a}\}$ and $B = \{b, \bar{b}\}$ in this particular order. Assuming the current state is $\{b\}$, $\{B\}$

is the expanded set of target(s), and $\{\}$ is the *unexpanded* set. So $\{b\}$ has no child state. As a result, different branches of the search tree may have different numbers of layers, but all of the states with the same cardinality appear in the same layer.

It is possible to use dynamic ordering to expand the targets that can most improve the GBF score first. However, it was shown in (Yuan and Hansen 2009) that a static ordering can actually make computing upper bounds and ultimately the search more efficient. We therefore simply ordered the targets according to their indices in this work.

An upper bound based on Markov blanket decomposition

For MRE inference, an upper bound of a state $\mathbf{x}$ should be greater than the GBF score of not only $\mathbf{x}$ but also all descendant states of $\mathbf{x}$. We can prune the whole subtree if the upper bound is less than the GBF of the current best solution.

We introduce the following novel upper bound for MRE inference. We first partition the evidences into a set of *exclusive* and *exhaustive* subsets, i.e., $\mathbf{E} = \cup_i \mathbf{E}_i$, such that $MB(\mathbf{E}_i) \subseteq \mathbf{M}$, by using a depth first search discussed later. These evidence subsets naturally decompose the whole network into overlapping subnetworks, each of which contains an evidence subset $\mathbf{E}_i$ and its Markov blanket $MB(\mathbf{E}_i)$. An upper bound on GBF is derived by multiplying the upper bounds on the belief update ratios calculated on the subnetworks.

We first derive an upper bound on the belief update ratio in Theorem 1.

Theorem 1. *Let $\mathbf{M} = \{X_1, X_2, \ldots, X_n\}$ be a set of targets, $\mathbf{e}$ be the evidence, and $MB(\mathbf{E}_i) \subseteq \mathbf{M}$ be the Markov blanket of the i^{th} subset of evidences $\mathbf{E}_i$. Then, for any subset $\emptyset \subset \mathbf{X} \subseteq \mathbf{M}$, the belief update ratio $r(\mathbf{x}; \mathbf{e})$ is upper bounded as follows.*

$$\max_{\mathbf{x}, \emptyset \subset \mathbf{X} \subseteq \mathbf{M}} r(\mathbf{x}; \mathbf{e}) \leq \left(\prod_i \max_{\mathbf{x}, \mathbf{X} = MB(\mathbf{E}_i)} r(\mathbf{x}; \mathbf{e}_i) \right) \cdot C, \quad (5)$$

where $C = \prod_i p(\mathbf{e}_i) / p(\mathbf{e})$.

Proof. From the formulation of $r(\mathbf{M}; \mathbf{e})$, we have

$$\begin{aligned} r(\mathbf{M}; \mathbf{e}) &= p(\mathbf{M}|\mathbf{e}) / p(\mathbf{M}) = p(\mathbf{e}|\mathbf{M}) / p(\mathbf{e}) \\ &= \prod_i p(\mathbf{e}_i | MB(\mathbf{E}_i)) / p(\mathbf{e}) \\ &= \prod_i \frac{p(MB(\mathbf{E}_i)|\mathbf{e}_i) p(\mathbf{e}_i)}{p(MB(\mathbf{E}_i))} / p(\mathbf{e}) \\ &= \left(\prod_i r(MB(\mathbf{E}_i); \mathbf{e}_i) p(\mathbf{e}_i) \right) / p(\mathbf{e}). \quad (6) \end{aligned}$$

The third equality is based on the property of Markov blankets.

Thus, we have:

$$r(\mathbf{M}; \mathbf{e}) = \left(\prod_i r(MB(\mathbf{E}_i); \mathbf{e}_i) \right) \cdot C, \quad (7)$$

where $C = \prod_i p(\mathbf{e}_i) / p(\mathbf{e})$.

From Equation 7, we have the following upper bound on $r(\mathbf{m}; \mathbf{e})$,

$$\max_{\mathbf{m}} r(\mathbf{m}; \mathbf{e}) \leq \left(\prod_i \max_{\mathbf{x}, \mathbf{X} = MB(\mathbf{E}_i)} r(\mathbf{x}; \mathbf{e}_i) \right) \cdot C. \quad (8)$$

For any $\emptyset \subset \mathbf{X} \subseteq \mathbf{M}$, $\mathbf{M} = \mathbf{X} \cup \bar{\mathbf{X}}$, let $\mathbf{S}_i = MB(\mathbf{E}_i) \cap \mathbf{X}$, $\mathbf{T}_i = MB(\mathbf{E}_i) \cap \bar{\mathbf{X}}$, $\mathbf{U} = \{i : \mathbf{T}_i \neq \emptyset\}$, and $\mathbf{V} = \{i : \mathbf{T}_i = \emptyset\}$, we have:

$$\begin{aligned} p(\mathbf{e}|\mathbf{X}) &= \sum_{\bar{\mathbf{X}}} p(\mathbf{e}|\mathbf{M}) p(\bar{\mathbf{X}}|\mathbf{X}) \\ &= \sum_{\bar{\mathbf{X}}} \left(\prod_{i \in \mathbf{U}} p(\mathbf{e}_i | MB(\mathbf{E}_i)) \right) p(\bar{\mathbf{X}}|\mathbf{X}) \cdot \\ &\quad \prod_{i \in \mathbf{V}} p(\mathbf{e}_i | MB(\mathbf{E}_i)) \\ &\leq \left(\prod_{i \in \mathbf{U}} \max_{\mathbf{z}, \mathbf{Z} = \mathbf{T}_i} p(\mathbf{e}_i | \mathbf{S}_i \cup \mathbf{z}) \right) \cdot \\ &\quad \left(\prod_{i \in \mathbf{V}} p(\mathbf{e}_i | MB(\mathbf{E}_i)) \right). \quad (9) \end{aligned}$$

Since $p(\mathbf{e}|\mathbf{X}) = r(\mathbf{X}; \mathbf{e}) p(\mathbf{e})$, we have the following upper bound on $r(\mathbf{x}; \mathbf{e})$:

$$\begin{aligned} \max_{\mathbf{x}} r(\mathbf{x}; \mathbf{e}) &\leq \left(\prod_{i \in \mathbf{U}} \max_{\mathbf{s}_i} \max_{\mathbf{z}, \mathbf{Z} = \mathbf{T}_i} r(\mathbf{z} \cup \mathbf{s}_i; \mathbf{e}_i) \right) \cdot \\ &\quad \left(\prod_{i \in \mathbf{V}} \max_{\mathbf{x}, \mathbf{X} = MB(\mathbf{E}_i)} r(\mathbf{x}; \mathbf{e}_i) \right) \cdot C. \quad (10) \end{aligned}$$

Thus, for any $\emptyset \subset \mathbf{X} \subseteq \mathbf{M}$, we obtain the final upper bound by combining Equations 8 and 10:

$$\max_{\mathbf{x}, \emptyset \subset \mathbf{X} \subseteq \mathbf{M}} r(\mathbf{x}; \mathbf{e}) \leq \left(\prod_i \max_{\mathbf{x}, \mathbf{X} = MB(\mathbf{E}_i)} r(\mathbf{x}; \mathbf{e}_i) \right) \cdot C.$$

□

In MRE inference, the evidence $\mathbf{e}$ is given, thus C is a constant. Theorem 1 assumes that the expanded target set $\mathbf{Y} = \emptyset$, which is true at the beginning of search. During the search when $\mathbf{Y} \neq \emptyset$, we have the following corollary to derive the upper bound on the belief update ratio:

Corollary 2. *Let $\mathbf{M} = \{X_1, X_2, \ldots, X_n\}$ be a set of targets, $\mathbf{e}$ be the evidence, $MB(\mathbf{E}_i) \subseteq \mathbf{M}$ be the Markov blanket of the i^{th} subset of evidences $\mathbf{E}_i$. Let $\mathbf{X}$ and $\mathbf{Y}$ be the unexpanded and expanded target sets. Let $\mathbf{T}_i = MB(\mathbf{E}_i) \cap \mathbf{Y}$, and $MB(\mathbf{E}_i) = \mathbf{T}_i \cup \bar{\mathbf{T}}_i$. Then, for any subset $\emptyset \subset \mathbf{Z} \subseteq \mathbf{X}$, the belief update ratio $r(\mathbf{z} \cup \mathbf{y}; \mathbf{e})$ is upper bounded as follows.*

$$\max_{\mathbf{z}, \emptyset \subset \mathbf{Z} \subseteq \mathbf{X}} r(\mathbf{z} \cup \mathbf{y}; \mathbf{e}) \leq \left(\prod_i \max_{\mathbf{z}, \mathbf{Z} = \bar{\mathbf{T}}_i} r(\mathbf{z} \cup \mathbf{t}_i; \mathbf{e}_i) \right) \cdot C, \quad (11)$$

where $C = \prod_i p(\mathbf{e}_i) / p(\mathbf{e})$.

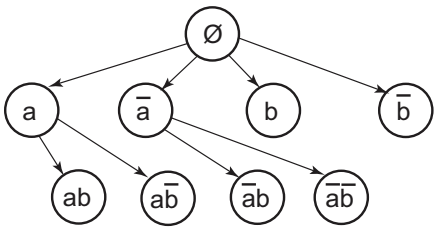

Figure 1: The search tree of an MRE problem with two targets, i.e., A and B.

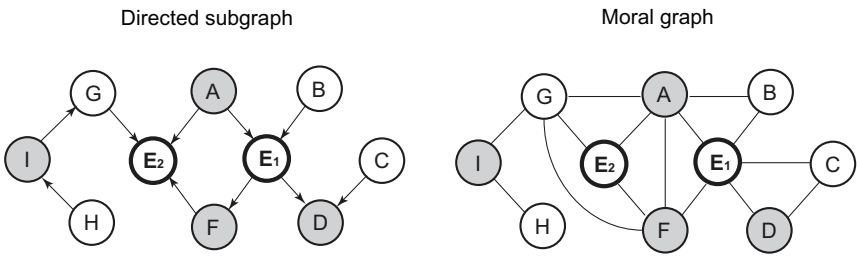

Figure 2: An example of compiling Markov blankets. E_1 and E_2 are evidence nodes. Gray nodes are targets. Others are auxiliary nodes. $MB(E_1)$ and $MB(E_2)$ share targets A and F.

Based on Corollary 2, we can derive the upper bound on GBF in Theorem 3:

Theorem 3. *Let $\mathbf{M} = \{X_1, X_2, \ldots, X_n\}$ be a set of targets, $\mathbf{e}$ be the evidence, $MB(\mathbf{E}_i) \subseteq \mathbf{M}$ is the Markov blanket of the i^{th} subset of evidences $\mathbf{E}_i$. Let $\mathbf{X}$ and $\mathbf{Y}$ be the unexpanded and expanded target sets. Let $\mathbf{T}_i = MB(\mathbf{E}_i) \cap \mathbf{Y}$, and $MB(\mathbf{E}_i) = \mathbf{T}_i \cup \bar{\mathbf{T}}_i$. Then, for any subset $\emptyset \subset \mathbf{Z} \subseteq \mathbf{X}$, the general Bayesian factor score $GBF(\mathbf{z} \cup \mathbf{y}; \mathbf{e})$ is upper bounded as follows.*

$$\begin{aligned} &\max_{\mathbf{z}, \emptyset \subset \mathbf{Z} \subseteq \mathbf{X}} GBF(\mathbf{z} \cup \mathbf{y}; \mathbf{e}) \\ \leq \; & 1 + \frac{\left(\prod_i \max_{\mathbf{z}, \mathbf{Z} = \bar{\mathbf{T}}_i} r(\mathbf{z} \cup \mathbf{t}_i; \mathbf{e}_i) \right) \cdot C - 1}{1 - p(\mathbf{y}|\mathbf{e})}, \end{aligned} \quad (12)$$

where $C = \prod_i p(\mathbf{e}_i)/p(\mathbf{e})$.

Proof. First, we formulate GBF using the belief update ratio as in Equation 4.

$$GBF(\mathbf{m}; \mathbf{e}) = 1 + \frac{r(\mathbf{m}; \mathbf{e}) - 1}{1 - p(\mathbf{m}|\mathbf{e})}.$$

For any subset $\emptyset \subset \mathbf{Z} \subseteq \mathbf{X}$, $p(\mathbf{z} \cup \mathbf{y}|\mathbf{e}) = p(\mathbf{z}|\mathbf{y}, \mathbf{e}) \cdot p(\mathbf{y}|\mathbf{e}) \leq p(\mathbf{y}|\mathbf{e})$. Thus we have:

$$\begin{aligned} &\max_{\mathbf{z}, \emptyset \subset \mathbf{Z} \subseteq \mathbf{X}} GBF(\mathbf{z} \cup \mathbf{y}; \mathbf{e}) \\ \leq \; & 1 + \frac{\max_{\mathbf{z}, \emptyset \subset \mathbf{Z} \subseteq \mathbf{X}} r(\mathbf{z} \cup \mathbf{y}; \mathbf{e}) - 1}{1 - p(\mathbf{y}|\mathbf{e})}. \end{aligned} \quad (13)$$

Then using Corollary 2, we obtain the following upper bound on GBF.

$$\begin{aligned} &\max_{\mathbf{z}, \emptyset \subset \mathbf{Z} \subseteq \mathbf{X}} GBF(\mathbf{z} \cup \mathbf{y}; \mathbf{e}) \\ \leq \; & 1 + \frac{\left(\prod_i \max_{\mathbf{z}, \mathbf{Z} = \bar{\mathbf{T}}_i} r(\mathbf{z} \cup \mathbf{t}_i; \mathbf{e}_i) \right) \cdot C - 1}{1 - p(\mathbf{y}|\mathbf{e})}, \end{aligned}$$

where $C = \prod_i p(\mathbf{e}_i)/p(\mathbf{e})$. □

Using Equation 12, we can bound all the descendant solutions of the current state $\mathbf{y}$. Equation 12 shows that in an MRE problem (left), we need to search all of the subsets of targets $\mathbf{M}$ to find the best solution. However, to calculate an upper bound (right), we only need to search a fixed target set $MB(\mathbf{E}_i)$ of each subnetwork, which usually has a small size and is easy to handle.

Compiling Markov blankets

The above theorems are based on factorizing the conditional joint distribution $p(\mathbf{e}|\mathbf{M})$ into the product of a set of conditional distributions $\prod_i p(\mathbf{e}_i|MB(\mathbf{E}_i))$. Thus partitioning the whole evidence set $\mathbf{E}$ into the subsets $\cup_i \mathbf{E}_i$ and compiling their Markov blankets is a key part of the proposed method. The Markov blanket of a single node includes its parents, children, and children's other parents. In MRE problems, some of these nodes may be auxiliary nodes. To derive the upper bound, however, we need to find the Markov blankets containing only targets. Thus the standard definition of Markov blanket cannot be directly used in our problem.

In the proposed method, we first generate the smallest ancestral set containing the target set $\mathbf{M}$ and the evidence set $\mathbf{E}$, i.e., $An(\mathbf{M} \cup \mathbf{E})$. Then we compile a moral graph $(G_{An(\mathbf{M} \cup \mathbf{E})})^m$. Figure 2 illustrates an example moral graph with two evidence nodes. Using Lemma 1, for each evidence subset $\mathbf{E}_i$, we need to find a set of targets which separates $\mathbf{E}_i$ from other evidence nodes and targets in $(G_{An(\mathbf{M} \cup \mathbf{E})})^m$. This is achieved by doing a depth first graph traversal starting from each evidence node if it has not been visited before. There are three scenarios when a node is being visited.

Case 1: When an evidence node is visited, we add the evidence to the current evidence subset $\mathbf{E}_i$, mark it as visited, and continue to visit its unmarked neighbors.

Case 2: When a target is visited, we add the target to $MB(\mathbf{E}_i)$ and mark it as visited.

Case 3: When an auxiliary node is visited, we mark it as visited and continue to visit its unmarked neighbors.

When restarting the search on a new evidence node, we unmark all targets and auxiliary nodes, because the same targets may occur in different Markov blankets as shown in Figure 2. The algorithm stops when all evidence nodes have been visited. The depth first search algorithm automatically partitions the whole evidence set into subsets $\cup_i \mathbf{E}_i$ and finds the Markov blanket $MB(\mathbf{E}_i)$ of each subset.

Merging Markov blankets

When computing the upper bound, we maximize each belief update ratio $r(MB(\mathbf{E}_i); \mathbf{e}_i)$ according to $MB(\mathbf{E}_i)$, independently. Thus the common targets of two different Markov blankets $MB(\mathbf{E}_i)$ and $MB(\mathbf{E}_j)$ may be set to inconsistent values based on the maximization. Too much inconsistency may result in a loose bound. We can tighten the upper bound by merging Markov blankets that share targets. On the other hand, if the number of targets in an individual Markov blanket is too large, it will make calculating the upper bound inefficient. We propose to merge Markov blankets to reduce the number of duplicates of a target under the constraint that the number of targets in a Markov blanket cannot exceed K.

We can use an undirected graph to represent the problem of merging Markov blankets. The nodes in the graph denote Markov blankets. If two Markov blankets share targets, there is an edge between the two corresponding nodes. The weight of edge is the number of targets shared by the two Markov blankets. This formulation translates the problem of merging Markov blankets into a graph partition problem. More specifically, the merging problem can be addressed by recursively solving the minimum bisection problem (Feige and Krauthgamer 2002), which partitions the vertices into two equal halves so as to minimize the sum of weights of the edges between the two partitions, on the undirected graph.

Minimum bisection problem is an NP-hard problem, however. We cannot afford to spend too much on computing the upper bound. We therefore use a hierarchical clustering-like greedy algorithm for merging the Markov blankets. We first merge any two Markov blankets if one of them covers the other. Then for all the remaining Markov blankets, we each time merge two Markov blankets that share the most number of targets as long as the number of targets in the resulting Markov blanket does not exceed K. The algorithm finishes when no Markov blankets can be merged.

Breadth-first branch-and-bound algorithm

In MRE inference, all of the search nodes are potential solutions except the root node. We can choose from a variety of search methods to explore the search tree, e.g., depth-first search, best-first search, and breadth-first search. We choose breadth-first search for two reasons. First, it is convenient to identify the convergence of MRE by monitoring the number of states in the search queue. Usually, the size of the search queue first increases and then decreases in the BFBnB algorithm. Second, since MRE prunes away independent and less relevant targets, usually the number of targets in the optimal solution is not large. Thus breadth-first search may reach the optimal solutions faster than other search strategies. The breadth-first search costs more memory to store the unexpanded states in a layer, however.

Before the breadth-first search, we calculate the belief update ratios of all configurations of $MB(\mathbf{E}_i)$ and store them in a belief table. Then the breadth-first search algorithm explores the search tree layer by layer while keeping track of the highest-scoring state. For each state S, we search for a configuration of each Markov blanket that is consistent with the expanded targets $\mathbf{t}_i$ and has the highest belief update ratio. We then calculate the upper bound of S using Theorem 3. We prune S if the upper bound is less than the current best GBF.

Networks	Nodes	Leaves	States	Arcs
Alarm	37	11	2.84	46
Carpo	61	43	2.23	74
Hepar	70	41	2.31	123
Insurance	27	6	3.30	52
Emdec6h	168	117	2.00	261
Tcc4g	105	69	2.00	193

Table 1: Benchmark diagnostic Bayesian networks used to evaluate the proposed algorithm.

Experiments

We tested the proposed algorithm on six benchmark diagnostic Bayesian networks listed in Table 1, i.e., Alarm (Ala), Carpo (Car), Hepar (Hep), Insurance (Ins), Emdec6h (Emd), and Tcc4g (Tcc) (Beinlich et al. 1989; Binder et al. 1997; Onisko 2003). Among them, Alarm, Carpo, Hepar, and Insurance are small networks with fewer than 100 nodes. Emdec6h and Tcc4g are larger networks with more than 100 nodes. We also listed the number of leaf nodes (Leaves), the average number of node states (States), and the number of arcs (Arcs) of the Bayesian networks in Table 1. The experiments were performed on a 2.67GHz Intel Xeon CPU E7 with 512G RAM running a 3.7.10 Linux kernel.

Experimental design

Since the proposed method is the first nontrivial exact MRE algorithm, we had to use a naive Breadth-First Brute-Force search algorithm (BFBF) as the baseline; basically BFBF is BFBnB with the bound set to be infinity. We also included the results of tabu search to indicate the difficulty of MRE problems. In BFBnB, we set the maximum number of targets in a Markov blanket K to be 18. In tabu search, we set the number of search steps since the last improvement L and the number of search steps M according to different network settings. For the networks with 12 targets, we set L to be 20 and M to be $\{400, 800, 1600, 3200, 6400\}$. For the networks with about 20 targets, we set L to be 80 and M to be $\{12800, 25600, 51200\}$. To evaluate search performance, we compared the solutions of tabu search and BFBnB, and counted the number of test cases on which tabu search achieved optimal solutions.

BFBF search can only solve test cases with fewer than 15 targets. To compare to BFBF, we randomly generated five test settings of each network, each setting consisting of all leaf nodes as evidence, 12 of the remaining nodes as targets, and others as auxiliary nodes. Then for each setting, we randomly generated 20 configurations of evidence (test cases) by sampling from the prior distributions of the networks.

We also evaluated BFBnB on test cases with about 20 targets. For each network, we generated five test settings and 20 test cases for each setting. In the four smaller networks, we set all the leaf nodes as evidence, 80% of the remaining nodes as targets, and others as auxiliary nodes. In the

sec	Ala	Car	Hep	Ins	Emd	Tcc
BFBF	1.7e3	66.1	270	1.6e5	212	162
BFBnB	17.0	3.5	14.6	1.6e3	15.1	5.6
T6400	92	100	93	81	95	100
	17.3	25.4	44.8	26.3	89.3	67.7
T3200	85	100	91	77	95	100
	9.1	13.2	22.6	13.7	45.0	36.1
T1600	79	100	86	74	95	100
	4.7	6.9	11.5	7.0	23.2	19.2
T800	76	98	81	73	95	100
	2.4	3.6	5.9	3.6	11.8	10.2
T400	72	98	81	73	95	100
	1.2	1.8	3.0	1.8	6.1	5.2

Table 2: Comparison of BFBnB, BFBF, and tabu on running time and accuracy on Bayesian networks with 12 targets.

two larger networks, we set all the leaf nodes as evidence, 20 of the remaining nodes as targets, and others as auxiliary nodes.

Evaluation of BFBnB on networks with 12 targets

In Table 2, we compared the proposed BFBnB algorithm with BFBF and tabu search on the test cases with 12 targets. For tabu search, we listed the accuracy (top) and running time (bottom) at different numbers of search steps. Both BFBnB and BFBF were able to solve the test cases exactly. BFBnB is shown to be significantly faster than BFBF because of the pruning by the upper bound. T400 is the fastest algorithm with the worst accuracy. With the increase of M, the running time of tabu search increased significantly. However, in most of the networks, tabu search could not solve all the test cases optimally, even using more running time than BFBnB.

To compare BFBnB and BFBF in detail, we computed the average running time of the individual test cases for the algorithms, and plotted the difference between the logarithmic running time of BFBnB and BFBF on each network as a point in Figure 3. We also drew the contour lines to highlight the differences. For example, the contour line marked by -3 contains the points on which BFBnB is 1000 times faster than BFBF. The results showed that although the averaged running time may change significantly, the ratio of running time between BFBF and BFBnB is relatively stable. BFBnB is roughly 10 to 100 times faster than BFBF.

Evaluation of BFBnB on networks with more targets

We also evaluated BFBnB on the test cases with 15 to 20 targets. We listed the number of targets of each network in the first row of Table 3. It is not surprising that BFBnB took longer to finish comparing with the results in Table 2, but the pruning of the upper bound slowed down the exponential growth of the running time significantly. Also, we can see that the running time of BFBnB depends on not only the number of targets, but also the tightness of upper bound and the Bayesian network structures, which control the number

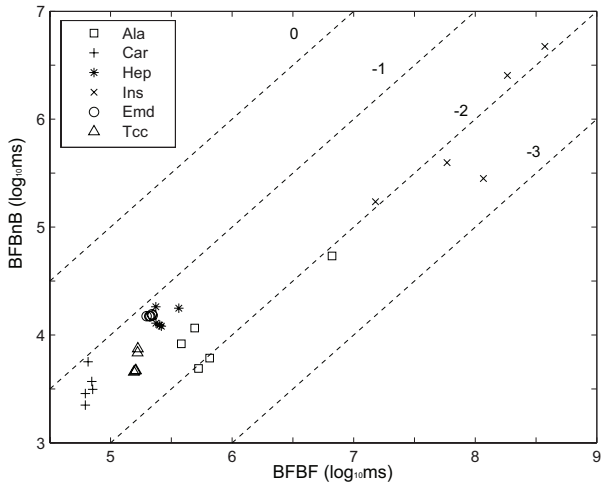

Figure 3: Distributions of logarithm running time pairs of BFBnB and BFBF on Bayesian networks with 12 targets.

min	Ala	Car	Hep	Ins	Emd	Tcc
	20	15	22	17	20	20
BFBnB	173	0.98	478	43.91	225	8.10
T51200	60	100	72	76	86	100
	4.06	5.55	17.12	0.96	34.14	30.92
T25600	60	100	65	76	86	100
	2.40	2.89	9.38	0.76	17.48	15.89
T12800	60	100	63	76	86	100
	1.42	1.50	5.15	0.59	8.93	8.10

Table 3: Comparison of BFBnB and tabu on running time and accuracy on Bayesian networks with 15 to 20 targets.

of pruned states and the size of belief table in each Markov blanket $MB(\mathbf{E}_i)$, respectively. Increasing M from 12800 to 51200 is not helpful in preventing tabu search from getting stuck in the local optima. Moreover, the performance of tabu search varies greatly on different networks.

Discussions and Conclusions

The main contribution of this paper is a BFBnB algorithm for solving MRE exactly using an upper bound based on Markov blanket decomposition. This is the first exact algorithm for solving MRE in Bayesian networks. The proposed upper bound can be calculated efficiently for two reasons. First, each Markov blanket $MB(\mathbf{E}_i)$ is usually much smaller than the whole target set $\mathbf{M}$. Second, in the original MRE problem, we need to search all the subsets of target set $\mathbf{M}$ to find the best solution. However, to calculate upper bounds, we only need to search on a fixed target set $MB(\mathbf{E}_i)$ of each subnetwork. The search space has been dramatically reduced. The experimental results show that BFBnB is significantly faster than the brute-force algorithm. BFBnB can solve MRE inference exactly in Bayesian networks which could not be solved previously.

Different from the brute-force algorithm, the search time of BFBnB is no longer mainly dependent on the size of

search space (i.e., the number of targets and the number of states of each target), but also on the tightness of the upper bound and the structures of Bayesian networks. For Bayesian networks with a large number of targets, an upper bound can be efficiently generated as long as the the number of targets in each Markov blanket is small. For the Markov blanket with a large number of targets, we can decompose it into subnetworks using the methods such as node splitting (Choi, Chavira, and Darwiche 2007). This is also one of our future research directions.

Acknowledgements This work was supported by NSF grants IIS-0953723, IIS-1219114, and a PSC-CUNY enhancement award.

References

Beinlich, I.; Suermondt, G.; Chavez, R.; and Cooper, G. 1989. The alarm monitoring system: A case study with two probabilistic inference techniques for belief networks. In *Proceedings of the 2nd European Conference on AI and Medicine*, 247–256.

Binder, J.; Koller, D.; Russell, S.; and Kanazawa, K. 1997. Adaptive probabilistic networks with hidden variables. *Machine Learning* 29:213–244.

Choi, A.; Chavira, M.; and Darwiche, A. 2007. Node splitting: A scheme for generating upper bounds in Bayesian networks. In *Proceedings of the 23rd Annual Conference on Uncertainty in Artificial Intelligence (UAI-07)*, 57–66.

Dechter, R., and Rish, I. 2003. Mini-buckets: A general scheme for bounded inference. *J. ACM* 50(2):107–153.

Feige, U., and Krauthgamer, R. 2002. A polylogarithmic approximation of the minimum bisection. *SIAM J. Comput.* 31(4):1090–1118.

Fitelson, B. 2007. Likelihoodism, bayesianism, and relational confirmation. *Synthese* 156:473–489.

Glover, F. 1990. Tabu search: A tutorial. *Interfaces* 20:74–94.

Good, I. J. 1985. Weight of evidence: A brief survey. *Bayesian statistics* 2:249–270.

Huang, J.; Chavira, M.; and Darwiche, A. 2006. Solving map exactly by searching on compiled arithmetic circuits. In *Proceedings of the 21st National Conference on Artificial intelligence*, volume 2, 1143–1148. AAAI Press.

Koller, D., and Friedman, N. 2009. *Probabilistic Graphical Models - Principles and Techniques*. MIT Press.

Lacave, C., and Díez, F. J. 2002. A review of explanation methods for Bayesian networks. *Knowledge Eng. Review* 17:107–127.

Lauritzen, S. L.; Dawid, A. P.; Larsen, B. N.; and Leimer., H.-G. 1990. Independence properties of directed markov fields. *Networks* 20(5):491–505.

Marinescu, R., and Dechter, R. 2009. And/or branch-and-bound search for combinatorial optimization in graphical models. *Artif. Intell.* 173(16-17):1457–1491.

Nielsen, U. H.; Pellet, J.-P.; and Elisseeff, A. 2008. Explanation trees for causal Bayesian networks. In *Proceedings of the 24th Annual Conference on Uncertainty in Artificial Intelligence (UAI-08)*, 427–434.

Onisko, A. 2003. *Probabilistic Causal Models in Medicine: Application to Diagnosis of Liver Disorders*. Ph.D. Dissertation, Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Science.

Pacer, M.; Lombrozo, T.; Griffiths, T.; Williams, J.; and Chen, X. 2013. Evaluating computational models of explanation using human judgments. In *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, 498–507.

Park, J. D., and Darwiche, A. 2001. Approximating map using local search. In *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01)*, 403–410.

Park, J. D., and Darwiche, A. 2003. Solving map exactly using systematic search. In *Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)*, 459–468.

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

Yuan, C., and Hansen, E. A. 2009. Efficient computation of jointree bounds for systematic MAP search. In *Proceedings of 21st International Joint Conference on Artificial Intelligence (IJCAI-09)*, 1982–1989.

Yuan, C., and Lu, T.-C. 2008. A general framework for generating multivariate explanations in bayesian networks. In *Proceedings of the 23rd National Conference on Artificial intelligence*, volume 2, 1119–1124. AAAI Press.

Yuan, C.; Liu, X.; Lu, T.-C.; and Lim, H. 2009. Most relevant explanation: Properties, algorithms, and evaluations. In *Proceedings of 25th Annual Conference on Uncertainty in Artificial Intelligence (UAI-09)*, 631–638.

Yuan, C.; Lim, H.; and Littman, M. L. 2011. Most relevant explanation: computational complexity and approximation methods. *Ann. Math. Artif. Intell.* 61:159–183.

Yuan, C.; Lim, H.; and Lu, T.-C. 2011. Most relevant explanation in Bayesian networks. *J. Artif. Intell. Res.* 42:309–352.