Abstract

This master thesis aims to propose propagation algorithms in singly connected causal belief networks where conditional beliefs are defined per edge by taking advantages of the representative power of the belief function theory. To ensure this propagation, we have first defined a new way to represent conditional distributions in the belief augmented network where conditional distributions are defined for each parent node (i.e., given the intervention, the DO node, and given the initial causes). Then, we have proposed a "centralized" propagation algorithm allowing the computation of the effects of both observations and interventions using the belief graph mutilation or the graph augmentation method. In the case of several observations and interventions, our second proposed "up-down" algorithm is more suitable since each node is visited at most twice regardless of the number of observed nodes.

Key words: belief causal networks, singly connected, belief function theory, conditional distributions, intervention, "centralized" propagation algorithm, "up-down" propagation algorithm.

Résumé

Ce rapport de mastre a pour but de proposer les algorithmes de propagation dans les réseaux causaux crédibilistes simplement connectés où les croyances conditionnelles sont définies par arc en tirant avantage de la théorie des fonctions de croyance. Pour assurer cette propagation, nous avons tout d'abord présenté une nouvelle façon pour définir les distributions conditionnelles dans le réseau augmenté qui sont définies pour chaque noeud parent (compte tenu de l'intervention, le noeud DO, et compte tenu des causes initiales). Ensuite, nous avons proposé un algorithme "centralisé" de propagation permettant le calcul des effets des observations et des interventions en utilisant soit la méthode de mutilation du graphe ou celle de l'augmentation du graphe. Dans le cas de plusieurs observations et interventions, notre second algorithme proposé "up-down" est plus approprié puisque chaque noeud est visité au maximum deux fois, indépendamment du nombre de noeuds observés.

Mots clés: réseaux évidentiels causaux crédibilistes, réseaux simplement connectés, théorie des fonctions de croyance, distributions conditionnelles, intervention, algorithme de propagation "centralisé", algorithme de propagation "up-down".

Acknowledgements

I would like to express my sincere gratitude to my advisor Professor Zied Elouedi for allowing me to conduct this research under his guidance and for his useful comments through the learning process of this master thesis.

My thanks must also go to my co-advisor Dr.Imen Boukhris for her patience, motivation, immense knowledge and help in all the time of research and writing of this thesis.

Special thanks to Wafa Laamari from LARODEC for her support.

I would like also to thank the jury for agreeing to evaluate my work.

I would like to thank my family: my parents for giving birth to me at the first place and supporting me spiritually throughout my life. To them I dedicate this master thesis.

Last but not the least, I would like to express my gratitude to all those who gave me the possibility to complete this work.

Contents

List of Figures

List of Tables

General Introduction

The notion of causality is a crucial concept in Artificial Intelligence when it comes to describe, interpret and analyze information and phenomena of our environment. Researchers in the AI field are interested in the problems arising from the modeling of causality motivated by the fact that it is important to provide the systems of inference or decision-making with explanation capacity for an operator or human user. Besides, it enables to anticipate the dynamics of events when the system is evolving using interventions which are external actions that force target variables to have specific values.

Graphical models are compact representations of uncertainty distributions. They are increasingly popular for reasoning under uncertainty due to their simplicity, their ability to easily express the human reasoning. Moreover, their capacity of representing and handling conditional independence relationships allows an efficient data management.

Bayesian networks (Pearl, 1998) are popular within the AI community. A probability distribution can be represented by several equivalent Bayesian networks since they induce the same joint distribution but only one of these networks follows the causal process which is the causal Bayesian network. However, probability distribution does not distinguish between equiprobability and ignorance situations. To tackle this problem, some authors (Xu & Smets, 1994), (Ben Yaghlane et al., 2003), (Simon et al., 2008), (Boukhris, Benferhat, & Elouedi, 2011) have proposed networks under the belief function framework which is an appropriate tool to express beliefs in a flexible way. These networks allow to compute the effect of passive observations of a system's spontaneous behavior.

Only the network proposed by (Boukhris, Benferhat, & Elouedi, 2011) allows to formalize imperfect causal knowledge. In fact, arcs do not only represent dependencies but also cause/effect relationships. These causal belief networks play an important role for the achievement of a coherent causal analysis.

Causal analysis is a more informative analysis. It allows not only to study the plausibility of an event under conditions of static experiments, but also to anticipate the dynamics of events when these conditions are changing. It means the capability to compute the effects of external actions (also called interventions) on the system.

While conditioning is used to compute the effect of observations, the "do" operator is used as a tool to compute the effect of interventions on belief causal networks. Handling interventions and computing their effects on the system can be done by making changes on the structure of the belief causal network.

To ensure the belief causal inference, we have to compute the effect of interventions and observations with two different equivalent ways, i.e., mutilating the graph by deleting the edges pointing to the node concerned by the action or using the augmented graph method by adding a new fictive variable, the variable "DO", as a new parent node of the variable A_i concerned by the intervention. This latter can take values in $do(x)$, $x \subseteq \{\Theta_{A_i} \cup \{\text{nothing}\}\}$. do(nothing) means that there are no actions on the variable concerned by the action and therefore represent the case of observations.

Existing algorithms only deal with the propagation of observational data in belief networks (Ben Yaghlane & Mellouli, 2008), (Xu & Smets, 1996), (Simon et al., 2008). Unfortunately, there is not any algorithm for belief causal inference, i.e., handling observational and interventional data. This have motivated us to propose algorithms for propagation in belief causal networks and to investigate causal inference in singly connected graphs under a belief function framework.

Belief inference algorithms are based on two rules proposed by (Smets, 1993a) called the Disjunctive Rule of Combination (DRC) and the Generalized Bayesian Theorem (GBT). These two operations allow to reason with conditional distributions instead of joint ones.

Contributions

In this dissertation, we propose three contributions regarding causal inference in singly connected networks where conditional beliefs are defined per edge allowing the compute the effect of observations and interventions consisting on:

- Representing conditional distributions in the belief augmented network where conditional distributions are defined for each parent node (i.e., given the intervention (the DO node) and given the initial causes).
- Proposition of an algorithm extended from the algorithm proposed by (Ben Yaghlane & Mellouli, 2008) so-called "centralized" algorithm.
- Proposition of an algorithm under a belief function framework extended from the standard Pearl's algorithm so-called "up-down" algorithm.

Organization

This master report is organized in four chapters as follows:

In Chapter 1, we first recall the basic concepts of the belief functions theory to model uncertainty. Then, a survey of existing belief tools related to networks under the belief function framework is provided at the end of this chapter.

In Chapter 2, we give the necessary background regarding the basic concepts of causality and its types. After that, we expose the existing belief causal networks.

In Chapter 3, the distribution regarding the "DO" node and the initial causes in the augmented graph will be proposed. Then, we propose several algorithms to handle observations and interventions.

In Chapter 4, tools for implementing our algorithms proposed in Chapter 3 for causal inference in singly causal belief networks were presented. Then, we provide results regarding these algorithms.

Finally, we summarize the results achieved in this master report and present some lines for future works.

Chapter 1

Belief Function Theory

1.1 Introduction

The theory of belief functions also known as Dempster-Shafer theory (Gordon & Shortliffe, 1984) or evidence theory has been developed by Arthur P. Dempster (Dempster, 1967) and generalized by Glenn Shafer (Shafer, 1976). It represents a new approach to model and manage imprecise and uncertain information in artificial and computational intelligence applications and a powerful tool for the reasoning under uncertainty because of its flexibility.

This theory has been applied by numerous authors. The field of application of this theory is large, e.g: expert systems (Gordon & Shortliffe, 1984), classification (Elouedi et al., 2001) (Trabelsi et al., 2011), diagnosis (Smets, 1998), pattern recognition (Denœux & Zouhal, 2001), etc.

Since this theory was developed, many interpretations have been proposed: those based on probability theory (lower probability model (Walley, 1991), the Dempster's model (1967, 1968) and the theory of hints (Kohlas & Monney, 1995)) and another interpretation relying on non-probabilistic theory (the Transferable Belief Model (TBM) (Smets, 1988a, 1993b, 1988b)).

In this master thesis, we deal with the interpretation of the belief function theory as explained by the TBM which is a model for representing beliefs. Beliefs can be held at two levels:

1. A credal level where beliefs are quantified by belief functions.

2. A pignistic level where decisions are made and beliefs are quantified by probability functions called pignistic probabilities.

In this chapter, we are interested by all aspects related to belief networks (i.e., networks using belief function theory (Xu & Smets, 1994), (Simon et al., 2008), (Ben Yaghlane & Mellouli, 2008) especially in their representation.
The rest of this chapter is organized as follows: in Section 1.2, we first present an overview of some basic concepts of belief theory. Section 1.3 is dedicated to special belief functions. Then, in Section 1.4, several basic operations are detailed. After this, in section 1.5, we provide a survey of the existing belief tools related to networks under the belief function framework.

1.2 Belief Function Teory: Basic concepts

1.2.1 Frame of discernment

Let Θ be a finite non empty set including all the elementary events related to a given problem. These events are assumed to be exhaustive and mutually exclusive. Such set Θ is called the frame of discernment.

The power set of Θ , denoted by 2^Θ is the set containing all the possible subsets of Θ. It is defined as follows:

$$2^\Theta = \{A : A \subseteq \Theta\}$$

A is an event designating either an elementary event or a disjunction of events.
The empty set $\emptyset$ belongs to the power set of Θ and it corresponds to the impossible event.

Example 1.1 *A murder has been committed. Suppose that the frame of discernment related to this problem is defined as follows:*
$\Theta = \{John,\ Mary,\ Peter\}$

The power set of Θ *is:*
$2^\Theta = \{\emptyset, \{John\}, \{Mary\}, \{Peter\}, \{John, Mary\}, \{John, Peter\}, \{Peter, Mary\}, \Theta\}$

1.2.2 Basic belief assignment

The impact of a piece of evidence on the different subsets of the frame of discernment Θ is represented by so-called **basic belief assignment**(bba).
The basic belief assignment, denoted by m, is a mapping from 2^Θ to [0,1] such that any proposition is associated with a real number belonging to [0,1] where the sum overall subsets is equal to 1:

$$\sum_{A \subseteq \Theta} m(A) = 1 \qquad (1.1)$$

The value m(A), named basic belief mass (bbm) represents the degree of belief committed to the event A of Θ . It this value cannot event or be allocated to any strict subset of A.

Shafer (1976) has initially considered that Θ includes all possible events. Thus, events composing the frame of discernment are exhaustive. Such bba is called a normalized basic belief assignment:

$$\mathrm{m}(\emptyset)=0$$

Smets (1990) relaxes this condition and considers that the frame of discernment may not be exhaustive. Therefore, m($\emptyset$) represents the part of belief supporting that none of the events in Θ is true.

Example 1.2 *Assume* Θ = {*John, Mary, Peter*}
The bba related to a piece of evidence concerning the murderer is defined as follows:

m({John})=0.6;
m({John, Mary})=0.2;
m(Θ)=0.2;

For example, 0.6 represents the part of belief exactly supporting that the murderer is John.

The subsets A of Θ such m(A) is strictly positive, are named **the focal elements**.

The union of all focal elements of m are named **the core** and are defined as follows:

$$\varphi = \bigcup_{A:m(A)>0} A \tag{1.2}$$

Example 1.3 *let's continue with the Example 1.2, the subsets {John}, {John,Mary}, and Θ are the focal elements of the bba m.*

The core of this bba m is defined as follows:
$\varphi = \{John\} \cup \{John, Mary\} \cup \Theta = \Theta$

1.2.3 Belief function

The belief function, denoted bel and corresponding to a specific bba m, assignes to every subset A $\subseteq \Theta$ the sum of the masses of belief committed exactly to every subset of A and to every proper subsets of A (Shafer, 1976).

It represents the total belief that one commits to A without being committed to $\bar{A}$.

The belief function *bel* is defined as follows:

$$bel : 2^{\Theta} \rightarrow [0,1] \quad \text{such that :}$$

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \tag{1.3}$$

Shafer notes that the function bel is named a belief function if and only if it satisfies the following conditions :

$$\text{bel}(\emptyset)=0$$
$$\text{bel}(\Theta)=1$$

$$bel(A_1 \cup A_2 \cup ... \cup A_n) \geq \textstyle\sum_i bel(A_i) -$$

$$\sum_{i>j} bel(A_i \cap A_j) - ... - (-1)^n bel(A_1 \cap ... A_n) \tag{1.4}$$

Example 1.4 *The belief function bel corresponding to the bba m (see Example 1.2) is defined as follows:*

bel(∅)=0;
bel({John})=0.6;
bel({Mary})= bel({Peter})=bel({Mary, Peter})=0;
bel({John, Mary})=0.6+0.2=0.8;
bel({John, Peter})=0.6;
bel(Θ)=0.6+0.2+0.2=1;

For example, 0.6 is the total belief committed to the proposition {John, Peter}.

1.2.4 Plausibility function

The plausibility function *pl* quantified the maximum amount of belief that could be given to a subset A of Θ. It is computed the total of masses compatible with A.

$$pl : 2^{\Theta} \rightarrow [0,1] \;\; \text{such that:}$$

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \tag{1.5}$$

$$pl(A) = bel(\Theta) - bel(A) \tag{1.6}$$

$$pl(A) = \sum_{B \subseteq \Theta} m(B) - \sum_{B \subseteq \bar{A}} m(B) \tag{1.7}$$

Example 1.10 *The plausibility function pl corresponding to the bba m (see Example 1.2) is defined as follows:*

pl(∅)=0;
pl({John})=0.6+0.2+0.2=1;
pl({Mary})=0.2+0.2=0.4;
pl({Peter})=0.2;
pl({John,Mary})=0.6+0.2+0.2=1;
pl({John,Peter})=0.6+0.2+0.2=1;
pl({Mary,Peter})=0.2+0.2=0.4;
pl(Θ)=0.6+0.2+0.2=1;

For example, 0.4 represents the maximum degree of belief that the proposition {Mary} may have.

1.2.5 Commonality function

The commonality function q is useful for simplifying some computations. It may represent the total mass that is free to move to every element of A (Barnett, 1981). It is defined as follows:

$$q : 2^{\Theta} \to [0,1] \text{ such that:}$$

$$q(A) = \sum_{A \subseteq B} m(B) \tag{1.8}$$

$$\mathrm{q}(\emptyset) = 1$$

$$\mathrm{q}(\Theta)=\mathrm{m}(\Theta)$$

Examlpe 1.11 *The commonality function q corresponding to the bba m (see example 1.2) is defined as follows:*

q(∅)=1;
q({John})=0.6+0.2+0.2=1;
q({Mary})=0.2+0.2=0.4;
q({Peter})=0.2;
q({John,Mary})=0.2+0.2=0.4;
q({John,Peter})=q({Mary,Peter})=0.2;
q(Θ)=0.2;

1.3 Special belief functions

In the literature, several kinds of belief functions are proposed. Such functions are used to express particular situations related generally to uncertainty.

1.3.1 Vacuous belief function

The case where the normalized bba quantifies the state of total ignorance and Θ is the unique focal element. This is a normalized belief function defined such that:

$$\text{m }(\Theta) = 1 \text{ and } \text{m(A)} = 0 \text{ for A} \neq \Theta$$

Example 1.5 *Assume an expert was not able to detect the murderer. Hence, we get a state of total ignorance where the corresponding bba is defined as follows:*

m(Θ)=1 and m(A)=0 for A $\neq \Theta$

1.3.2 Categorical belief function

The case where the normalized bba has a unique focal element A different from the frame of discernment Θ. This is a normalized belief function defined as:

$$\text{m(A)=1 for some A} \subset \Theta \text{ and m(B)=0, for B} \subset \Theta\text{, B} \neq \Theta$$

Example 1.7 *Assume an expert was certain that the murderer is a man. So, the corresponding bba presents a categorical belief function defined as follows:*

m({John, Peter})=1;

1.3.3 Certain belief function

The case where the bba quantifies the total certainty. There is exactly one focal element that is a singleton.

$$\text{m }(\theta) = 1 \text{ for one particular element of } \Theta$$

Example 1.6 *Assume an expert affirms that the murderer is Peter. So, the corresponding bba presents a certain belief function defined as follows:*

m({Peter})=1;

1.3.4 Bayesian belief function

The case where all focal elements are singletons are singletons. It is a particular case of probabilities and is defined as follows:

$$\mathrm{m}(\theta) \succ 0 \text{ for some elements of } \Theta$$

Example 1.8 *Let's consider* Θ*=*{*John, Mary, Peter*}*.*
We get a piece of evidence expressed by the following bba m:

m({*John*}*)=0.3;*
m({*Mary*}*)=0.4;*
m({*Peter*}*)=0.3;*
m(Θ*)=0;*

1.3.5 Consonant belief function

The case where all the focal elements are nested. It is a special case of possibility theory (Dubois et al., 2001).

Example 1.9 *Let's consider the same bba defined in the Example 1.2:*

m({*John*}*)=0.6;*
m({*John,Mary*}*)=0.2;*
m(Θ*)=0.2;*

1.3.6 Simple support function

The case where at most one focal is different different from the frame of discernment Θ.

Example 1.9 *Let's consider the same bba defined in the Example 1.2. Assume we have a bba defined as follows:*

m({*Mary, Peter*}*)=0.7;*
m(Θ*)=0.3;*

m is called a simple support function where the focus is the proposition {*Mary,Peter*}.

1.3.7 Non-dogmatic and dogmatic belief functions

A belief function is said to be non-dogmatic if the frame of discernment is a focal element, i.e., m(Θ) $\succ$ 0.

A belief function is said to be dogmatic if and only if its bba m is defined as m(Θ)=0.

1.4 Belief Function Teory: Basic operations

1.4.1 Combination

The belief function theory offers interesting tools for aggregating basic belief assignments. These beliefs provided by different and distinct sources can be combined either conjunctively or disjunctively.

The conjunctive rule of combination

Consider two distinct pieces of evidence on Θ represented by m_1 and m_2. The belief function that quantifies the combined impact of these two pieces of evidence is obtained through *the conjunctive rule of combination.* It is defined as follows (Smets, 1998):

$$(m_1 \textcircled{\cap} m_2)(A) = \sum_{B,C \subseteq \Theta, B \cap C = A} m_1(B) m_2(C) \tag{1.9}$$

This rule can also be written in terms of commonality functions as follows:

$$(q_1 \textcircled{\cap} q_2)(A) = q_1(A) \cdot q_2(A) \tag{1.10}$$

Dempster's rule is considered as the normalized conjunctive rule of combination and is defined as follows (Shafer, 1976, 1986):

$$m_1 \oplus m_2(A) = \begin{cases} m_1 \textcircled{\cap} m_2(A)/1 - m_1 \textcircled{\cap} m_2(\emptyset) & if \quad A \neq \emptyset, A \subseteq \Theta \\ 0 \quad otherwise \end{cases} \tag{1.11}$$

The $\oplus$ represents the conjunctive combination where the normalization is not performed.

The conjunctive rule of combination has the following properties:

- Commutative:
$$m_1 \oplus m_2 = m_2 \oplus m_1$$
- Asoociative:
$$(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$$
- Non-idempotent:
$$m_1 \oplus m_1 \neq m_1$$
- Neutral element: The neutral element within the conjunctive rule is the vacuous basic belief assignment representing the total ignorance.

$$m \oplus m_0 = m$$

Disjunctive rule of combination

Consider two distinct pieces of evidence on Θ represented by m_1 and m_2. Suppose an agent knows only that at least one of the pieces of evidence prevails without knowing which one. Then, the belief function that quantifies these two pieces is obtained through *the disjunctive rule of combination*, denoted by $\textcircled{\cup}$ and defined as follows (Smets, 1998):

$$(m_1 \textcircled{\cup} m_2)(A) = \sum_{B,C \subseteq \Theta, B \cup C = A} m_1(B) m_2(C) \tag{1.12}$$

1.4.2 Cylindrical extension and projection

Cylindrical extension

The cylindrical extension is an operation allowing the extension of a set from a low-dimensional domain to a higher-dimensional domain.

Let us consider two distinct frame of discernment Ω and Θ and A is a subset of Θ. The cylindrical extension of A to $\Theta \times \Omega$ is defined as follows:

$$A^{\uparrow \Theta \Omega} = A \times \Omega \tag{1.13}$$

Projection

The projection is an operation allowing the dropping of extra coordinates and the reduction of a set defined in multi-dimensional domain to a set defined in a lower-dimensional domain.

Let us consider two distinct frames of discernment Ω and Θ, and A is a subset of Θ. Projecting A on Ω is defined as follows:

$$A^{\downarrow \Omega} = \{\omega, \omega \in \Omega, A \cap \omega^{\uparrow \Theta \Omega} \neq \emptyset\} \tag{1.14}$$

1.4.3 Vacuous extension and marginalization

Vacuous extension

Vacuous extension is useful when new variables are added to the referential. It allows to express the marginal mass function m^{Θ} defined on Θ over the frame $\Theta \times \Omega$ as follows:

$$m^{\Theta \uparrow \Theta \Omega}(B) = m^{\Theta}(A) \;\; if \;\; B = A \times \Omega \;\; such \;\; that \;\; A \subseteq \Theta, B \subseteq \Theta \times \Omega \tag{1.15}$$

It corresponds to make a cylindrical extension of A to $\Theta \times \Omega$.

Marginalization

Given the product space $\Theta \times \Omega$ and a mass distribution defined on this product space. Marginalization allows to map over a subset of the product space by dropping the extra coordinates.

$$m^{\Theta\Omega\downarrow\Theta} = \sum_{C\subseteq\Theta\times\Omega, C^{\downarrow\Theta}=A} m^{\Theta\Omega}(C), A \subseteq \Theta \tag{1.16}$$

It corresponds to projecting C on Θ.

1.4.4 Ballooning extension

The ballooning extension (Ristic & Smets, 2005) (Ben Yaghlane & Mellouli, 2008) is useful when, after conditioning, an expert change his belief in the light of new information and he would reconstruct initial distribution.

The ballooning extension transforms a conditional belief function $m(A|\omega)$ defined on Θ for $\omega \in \Omega$ into a new belief function over $\Theta \times \Omega$. To get rid of conditioning, we have to compute the ballooning extension defined as:

$$m_{\omega}^{\Theta\Uparrow\Theta\Omega}(C) = \begin{cases} m^{\Theta}(A|\omega) \;\; if \;\; C = (A \times \omega \cup \Theta \times \bar{\omega}) \\ 0 \;\; otherwise \end{cases} \tag{1.17}$$

$$such \;\; that \;\; A = (C \cap \omega^{\Theta\Omega})^{\downarrow\Theta}$$

1.4.5 Dempster's rule of conditioning

Dempster's rule of conditioning is one of the natural ingredients of the transferable belief model (Smets, 1990). In addition, it allows us to update the knowledge of an expert (who allocates a mass to a proposition A) had in the light of new information that an event B $\subseteq \Theta$ is true.

$m(A|B)$ denotes the degree of belief of A in the context of B with A,B $\subseteq \Theta$. The Dempster's rule of conditioning is computed as follows:

$$m(A|B) = \begin{cases} K.\sum_{C\subseteq\bar{B}} m(A \cup C) \;\; if A \subseteq B, A \neq \emptyset \\ 0 \;\; if \;\; A \not\subseteq B \end{cases} \tag{1.18}$$

where K^{-1}=$1 - m(\emptyset)$. It is called the normalization factor.

Dempster's rule of conditioning is seen as a special case of Dempster's rule of combination.

The conjunctive rule of combination is expressed using the unormalized Dempster's rule of conditioning (Dubois & Prade, 1986).

$$f_1 \oplus f_2 = \sum_{B\subseteq\Theta} f_1(A|B) m_2(A) \;\; where \;\; f \in [m, bel, pl, q] \tag{1.19}$$

1.4.6 Disjunctive rule of combination (drc)

The belief function induced by the set of conditional belief functions can be conditioned on x $\subseteq \Theta_X$ and the result marginalized on Θ, corresponding to the so-called *Disjunctive rule of combination*, proposed by Smets (1993a). If we want to compute f($x|\theta$); f $\in$ [b, bel, pl, m] for any $\theta \in \Theta$ and x $\in \Theta_X$, we use the DRC which allows us to build functions as follows:

$$\forall \theta \in \Theta, \forall x \in X$$

$$b^X(x|\theta) = \prod_{\theta_i \in \theta} b^X(x|\theta_i) \tag{1.20}$$

where b is the implicability function such that :

$$b : 2^\Theta \rightarrow [0,1] :$$

$$b(A) = bel(A) + m(\emptyset) \quad such \quad that \quad A \subseteq \Theta \tag{1.21}$$

$$bel^X(x|\theta) = b^X(x|\theta) - b^X(\emptyset|\theta) \tag{1.22}$$

$$pl^X(x|\theta) = 1 - \prod_{\theta_i \in \theta} (1 - pl^X(x|\theta_i)) \tag{1.23}$$

$$m^X(x|\theta) = \sum_{(\cup_{i:\theta_i \in \theta} x_i) = x} \prod_{i:\theta_i \in \theta} m^X(x_i|\theta_i) \tag{1.24}$$

1.4.7 Generalized bayesian theorem (gbt)

Smets (1993a) has generalized the bayesian theorem within the transferable belief model framework.

Let us consider two distinct variables X and Y defined on the spaces Θ_X and Θ_Y respectively. Let X be the parent of Y. For any $x \subseteq \Theta_X$ and $y \subseteq \Theta_Y$, f($x|y$) represents the conditional belief function induced on y given x_i element of Θ_X where A is one of the belief function measures (m, bel, pl and q). It is computed using the generalized bayesian theorem (GBT).

This belief is a posterior belief function and it is defined as (Ristic & Smets, 2005):

$$\forall \theta \in \Theta, \forall x \in X$$

$$b_\Theta(\theta|x) = \prod_{\theta_i \in \bar{\theta}} b_X(\bar{x}|\theta_i) \tag{1.25}$$

$$bel_\Theta(\theta|x) = b_\Theta(\theta|x) - b_\Theta(\emptyset|x) \tag{1.26}$$

$$pl^\Theta(\theta|x) = 1 - \prod_{\theta_i \in \theta}(1 - pl^X(x|\theta_i)) \tag{1.27}$$

$$q^\Theta(\theta|x) = \prod_{\theta_i \in \theta} pl^X(x|\theta_i) \tag{1.28}$$

1.5 Discounting

Dealing with evidence expressed by experts requires to take into account the reliability of an expert by the discounting method defined as:

$$m^\alpha(A) = \begin{cases} (1-\alpha).m(A), \forall \;\; A \subset \Theta \\ \alpha + (1-\alpha).m(A), \;\; if \;\; A = \Theta \end{cases} \tag{1.29}$$

The discounting operation is controlled by a *discount rate* α taking values between 0 and 1.

- α=0 means that the expert is totally reliable.
- α=1 means that the expert is not reliable at all. His opinions have to be totally ignored.

1.6 Description of existing networks

Several graphical models (e.g., probabilistic Bayesian networks (Darwiche, 2009) (Jensen & Nielsen, 2007) (Pearl, 1988), possibilistic networks (Ben Amor, Benferhat, & Mellouli, 2003) (Benferhat & Smaoui, 2007), credal networks (Cozman, 2000), valuation networks (Shenoy, 1993), belief function networks (Ben Yaghlane & Mellouli, 2008) (Xu & Smets, 1996) are increasingly popular knowledge representations for reasoning under uncertainty.

In this section, we briefly recall networks formalized with the belief function theory named belief networks. We investigate the evidential network (EN) proposed by (Simon et al., 2008), the directed evidential networks namely the evidential networks with conditional belief functions (ENC) proposed by (Xu & Smets, 1994), the directed evidential networks (DEVN) presented by (Ben Yaghlane & Mellouli, 2008), and the belief network with conditional beliefs proposed by (Boukhris, Elouedi, & Benferhat, 2011a). These networks can be categorized according to the way conditional distributions are defined: those where conditional distribution are defined for all parents (EN), those where conditionals are defined per single parent (ENC, DEVN) and finally those were conditionals are defined per some parents (BNC).

1.6.1 Conditional distribution for all parents

Evidential Networks(EN)

The evidential network, proposed by (Simon et al., 2008), is a belief network that combines the Dempster-Shafer theory with Bayesian Network (BN).

An evidential network is defined on two levels:

- Qualitative level: represented by a DAG G=(V,E), where V represents the set of variables. Each variable A_i is associated with a finite set namely its frame of discernment Θ_{A_i}. E represents the set of edges that encode the dependencies among variables.
- Quantitative level: represented by the set of masses distributions associated to each node in the graph.
  - For each root node (i.e., node without parent nodes) having a frame of discernment Θ_{A_i}, an a priori mass distribution m^{A_i} has to be defined over the power set $2^{\Theta_{A_i}}$.
  - For other nodes, a conditional mass distribution $m^{A_i}(.|Pa(A_i)$ is specified for each value of A_i knowing the value of all the parents $Pa(A_i)$.

The goal of this network is the modeling of reliability in a compact and graphic form under epistemic uncertainty. This network was applied to illustrate complex systems, precisely the Oil Pipeline System.
The proposed network can be seen as a standard Bayesian Network integrating the evidence theory where nodes take the following states: {Up}, {Down}, {Up,Down}.

Computation of the global joint distribution

As for bayesian network, the global joint distrubution over the set of variables is unique and can be expressed as a product of all conditional beliefs as follows:

$$m^V = \prod_{i=1}^{n} m^{A_i}(.|Pa(A_i)) \tag{1.30}$$

1.6.2 Conditional distribution per single parents

Evidential Networks with conditional beliefs(ENC)

Evidential networks with conditional belief functions, called ENC, was originally proposed by (Smets, 1993a) for the propagation of beliefs. Then, they have been developped by (Xu & Smets, 1994). The goal of the proposed

network is to simplify knowledge acquisition and storage.
In these networks, conditional beliefs are defined in a different way from conditional probabilities in the Bayesian networks (BNs) and Evidential networks (EN) and are used to represent binary relations between node: each edge represents a conditional relation between the two nodes it connects.

Graphically, the network is represented by a directed acyclic graph (DAG), G=(V,E). Nodes represent random variables where each variable is associated with a finite set of all its possible values. If there is an edge from variable A_i to variable A_j, A_i is the parent of A_j.
In these networks, directed edges describe the conditional dependencies in the model and they are associated with beliefs where an edge represents a conditional relation between the two nodes it connects.
In Figure 1.1, the edges (X,Z) and (Y,Z) mean that we have $\{bel^{\Theta_X}(x_i) : x_i \in \Theta_X\}$ and $\{bel^{\Theta_Y}(y_i) : y_i \in \Theta_Y\}$, but not $\{bel^{\Theta_{XY}}(x_i, y_i) : x_i \in \Theta_X, y_i \in \Theta_Y\}$ as proposed in Bayesian network.

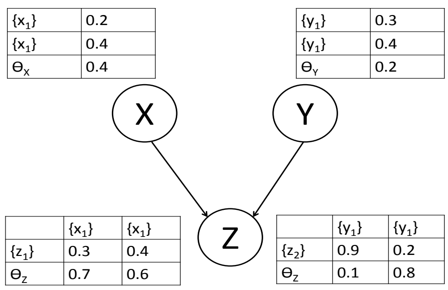

Figure 1.1: The directed evidential network

In an ENC, if a conditional belief is given in the context of more than one parent, these nodes should be merged into one node.

Example 1.12 *Let us consider a DAG with three nodes X, Y and Z where X and Y are parents of Z with the following conditional distribution* $m^Z(.|x_1, y_1)$ *where* $x_1 \in \Theta_X, y_1 \in \Theta_Y$ *. In this case, the node X and Y should be merged into one node (see Figure 1.2).*

Applying this fusion process does not allow to recover the a priori beliefs about parent nodes. So, this network leads to a loss of information.

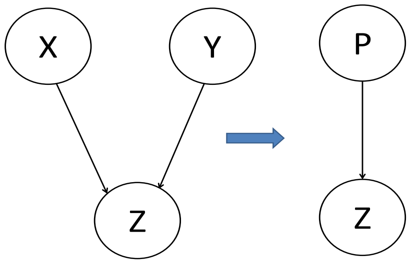

Figure 1.2: Evidential network with conditional beliefs

Directed evidential networks with conditional belief functions (DEVN)

The so-called Directed Evidential Networks with conditional beliefs are proposed by (DEVN) (Ben Yaghlane & Mellouli, 2008) in order to generalize the ENC where conditional beliefs are defined in a different manner from conditional probabilities in the BN. Note that beliefs can be defined for all parent nodes. Therefore, there is no need to merge nodes into one node.

This model can be defined on two levels:

- Qualitative level: represented by a direct acyclic graph (DAG), G=(V,E) in which the nodes represent variables, and directed arcs describe the conditional dependence relations embedded in the model.
- Quantitative level: represented by the set of beliefs (bel) associated to each edge in the graph. An edge represents a conditional relation between the two nodes it connects.

Computation of the global joint distribution

Given all the a priori and conditional belief functions, the joint distribution (Ben Yaghlane et al., 2003) relative to the set of variables $(A_1 \ldots A_n)$ is obtained by combining the joint distribution of each node using the following belief chain rule (Ben Yaghlane et al., 2003). This chain rule is computed either for ENC or DEVN with belief function distributions (bel) as follows:

$$bel^{A_1 \ldots A_n} = \textcircled{\cap}_{i=1 \ldots n} (\textcircled{\cap}_{\omega \in Pa(A_i)} bel^{A_i}(.|\omega)^{\uparrow A_i \times Pa(A_i)}) \tag{1.31}$$

1.6.3 Conditional distribution per some parents

Belief network with conditional beliefs (BNC)

The so-called belief network with conditional beliefs (BNC) (Boukhris, Elouedi, & Benferhat, 2011a) is a flexible network since conditional distribution may be defined given one parent or more than one without necessarily have to define all the parents like in DEVN and in EN or without need to have binary relations between nodes like in ENC.

This model is defined on two levels:

- Qualitative level: represented by a DAG G=(V,E) in which the nodes represent variables, and directed arcs describe the conditional dependence relations embedded in the model. represented by the set of bbas associated to each node in the graph.
- Quantitative level: dependence relations are expressed by the set of bbas associated to each node in the context either of one or more than one parent.
  - for each root node A_i having a frame of discernment Θ_{A_i}, an a priori mass distribution m has to be defined over the power set $2^{\Theta_{A_i}}$, such that:

$$\sum_{sub_{ik} \subseteq \Theta_{A_i}} m^{A_i}(sub_{ik}) = 1, k = 2, \ldots, 2^{\Theta_{A_i}}$$

  - for other nodes, a conditional belief mass distribution $m^{A_i}(.|Pa(A_i))$ is specified for each value of A_i knowing the focal sets of its parents defined by $Pa(A_i)$.

$$\sum_{sub_{ik} \subseteq \Theta_{A_i}} m^{A_i}(sub_{ik}|Pa(A_i)) = 1$$

Computation of the global joint distribution

To compute the global joint belief distribution, the ballooning extensions for the deconditionalization process are first computed for local conditional distributions. All deconditionalized beliefs are vacuously extended and aggregated.

In the general case, i.e., where beliefs are given per single parent, the computation of the global joint distribution is done in three steps:

1. For a conditional variable A_i:

   (a) For each subset of a single parent denoted by $Pa_j(A_i)$, compute the ballooning extension of $m^{A_i}\ (.|Pa_j(A_i))$ for the deconditionalization process:

   $$m^{A_i}(.|Pa_j(A_i))^{\Uparrow A_i \times PA_j(A_i)}$$

   (b) Combine the deconditionalized beliefs using the conjunctive rule of combination.

   $$\oplus_{Pa_i(A_i)} m^{A_i}(.|Pa_j(A_i))^{\Uparrow A_i \times PA_j(A_i)}$$

   where $PA_j(A_i)$ is a single parent of A_i and $Pa_j(A_i)$ is a subset from $PA_j(A_i)$.

2. Extend each node (root node and child node) to the universe of all the variables in the network by applying the vacuous extension.

   $$(\oplus_{Pa_i(A_i)} m^{A_i}(.|Pa_j(A_i))^{\Uparrow A_i \times PA_j(A_i)})^{\uparrow (A_1 \times \ldots \times A_n)}$$

3. Combine local joint distributions using the conjunctive rule of combination and thus get the following chain rule:

$$m^{(A_1 \ldots A_n)} = \oplus_{i=1 \ldots n} (\oplus_{Pa_i(A_i)} m^{A_i}(.|Pa_j(A_i))^{\Uparrow A_i \times PA_j(A_i)})^{\uparrow (A_1 \times \ldots \times A_n)} \tag{1.32}$$

1.7 Conclusion

In this chapter, we have presented the basic concepts and operations of the belief function theory to handle uncertainty. Then, we have explained that belief networks are very important to model uncertainties and to simplify knowledge acquisition.

The belief function theory will be used as a tool to formalize the imperfect causal knowledge which may be represented with a causal belief network. Thus, in the next chapter, we will represent the notion of causality.

Chapter 2

Causality modeling

2.1 Introduction

Causality is a crucial concept and plays an important role in many fields, from physics to medicine to Artificial Intelligence (AI). Researchers in the AI field were interested in the problems arising from the modeling of causality motivated by the fact that it is important to provide the systems of inference or decision-making with explanation capacity for an operator or human user (Dubois & Prade, 2003).

Causality amounts to determine what truly causes what and what it matters. It allows to describe, interpret and analyze information and events and it plays an important role in the expression of our perception of our environment. Besides, it enables to anticipate the dynamics of events when the system is evolving using interventions which are external actions that force a target variable to have a specific value (standard intervention) or more than one specific value (non-standard intervention). In fact, causal knowledge enables to predict future events and thus choose the right actions to achieve the goals.

Issues related to causality have been widely addressed and have been debated for many years. Rather recently researchers were interested in the problems arising from the modeling of causality (Shafer, 1996), (Pearl, 1998, 2000), (Halpern & Pearl, 2005), (Dubois & Prade, 2003), etc.

Causality is compactly represented with graphical models. On these causal networks, we can compute the simultaneous effect of observing the natural behavior of the system and external actions. Causal networks are appropriate tools to model causal knowledge which is usually uncertain. They provide information about the dynamics of the system under study.

While conditioning is used to compute the effect of observations, the "do" operator (Pearl, 2000) is used as a tool to represent interventions on causal networks. Thus, this operator is used to compute the impact of external action (Boukhris, Elouedi, & Benferhat, 2013). Handling interventions and computing their effects on the system can be done by making changes on the structure of the causal network. The two equivalent methods developed were called, graph mutilation and graph augmentation methods.

A standard intervention is a certain action which always succeeds to put its target at a precise value by making it completely independent of its original causes is a condition that is rarely achieved in real world applications. An intervention can be imperfect (uncertain or imprecise), this kind of intervention is called a non-standard intervention. Besides, if it takes place with a degree of belief, it can have imperfect consequences which means that it may not succeed to put its target into one specific value.

In this chapter, we present an overview of some definitions of causality in Section 2.2. In Section 2.3., we present the distinction between interventions and observations is detailed and also we present also standard and non-standard interventions. After this, we provide a survey of the existing belief causal networks respectively in Sections 2.4., 2.5 and 2.6.

2.2 Causality: Definitions

Researchers were interested in modeling of causality. Since there is no consensus regarding this concept, we will represent some definitions:

2.2.1 Causation vs association

An association is a relationship between two events. It allows to make them dependent but not necessarily causal. However, causality (also referred to as causation) is the relationship between an event (the cause) and a second event (the effect), where the second event is understood as a consequence of the first.

Association is a symmetric relation (i.e., A_1 associated to A_2 entails that A_2 is also associated to A_1). However, causation is an asymmetric relation (i.e., A_1 caused A_2 does not imply that A_2 caused A_1). However, a causal relation can be symmetric (i.e., A_1 causes A_2 and A_2 causes A_1).

Example 2.1 *Let us denote by H, the fact of having blond hair and by E the fact of having light-colored eyes. These two events are correlated. Moreover, the relation between them is a symmetric relation. However, this does not prove that having blond hair causes having light-colored eyes.*

Let us denote by S, the fact of smoking and by C the fact of having lung cancer. The two events are correlated and the their association involves causation. The relation between them is an asymmetric relation. In fact, smoking causes lung cancer but lung cancer does not cause smoking.

2.2.2 Counterfactuals

A counterfactual is a conditional statement which tells what would have happened if events other than the ones we are currently observing had happened. The basic idea of counterfactual theories of causation is that the meaning of causal claims can be explained in terms of counterfactual conditionals of the form "If A had not occurred, C would not have occurred". (Hume, 2006) (Pearl & Hopkins, 2007).

Counterfactuals are based on the existence of relations between possible worlds (interpretation) (i.e., the meaning of "If A had not occurred, C would not have occurred", there is at least one world in which A did not occurred and C did not occurred, that is closer to the actual world than other interpretations like A did not occurred and C occurred.

Example 2.2 *Let us consider this claim: "If I had the high school diploma with a very good average, I would have done medicine". This interpretation supposes that among the possible worlds where I had the high school diploma with a very good average, it exists at least on world where I would have done medicine which is closer than where I had the high school diploma with a very good average and have not done medicine.*

2.3 Observations vs Interventions

In the following, we present the difference between standard interventions and observations.

2.3.1 Observation

Observing is the action of seeing and monitoring of phenomena happened by themselves without any manipulation on the system. It can provide new information about the value of a variable and the statistical relations amongst events in a static world.

Observations of events allow to reason backwards diagnostically to infer their causes or to reason forward and predict future effects. They inform about other events that are directly or indirectly causally related to the observed event. Observing an event increases the probability of its causes and of its effects.

Example 2.3 *If someone has a high level of cholesterol, then you can make the diagnostic inference that he has probably followed an unhealthy diet (cause) and you can predict that his risk of contracting heart problems is relatively high (effect).*

Formally, observations are modeled by setting the event variables to the values that have been observed. The effect of other events conditional on the observed variable can be computed thanks to the structure of the causal model.

2.3.2 Intervention

Intervention is a crucial notion to insure an efficient causal analysis in the sense that it facilitates causality ascriptions. Intervening ((Pearl, 2000), (Spirtes, 2001)) (doing and the act of manipulating) is an external action that perturbs the spontaneous behavior of the system by forcing a variable to take a specific value (standard intervention) or more than one value (non-standard intervention). It allows to better anticipate the system evolution when such events occur in a dynamic world and is determined as coming from outside the system. It means that the natural behavior of an object is voluntary changed. Therefore, interventions allow the identification of elements in a sequence of events that are related in a causal way.

Observational data provide some information about the statistical relations among events. It means that events might be correlated without necessarily following a causal process. To tackle this problem, interventional data is used.

Unlike observations, interventions do not provide positive or negative diagnostic evidence about the causes of the event we intervened upon. In fact, interventions make the occurrence of events independent of their typical causes.

Handling non-standard interventions

In real world applications, assuming that an intervention is always perfect is not usually true. In fact, considering the intervention as a certain action which always succeeds to put its target at a precise value by making it completely independent of its original causes is a condition that is rarely achieved. Therefore, some works investigate non-standard interventions (Boukhris et al., 2012; Boukhris, Benferhat, & Elouedi, 2013);(Eberhardt & Scheines, 2007); (Korb et al., 2004); (Teng, 2012); (Woodward, 2003).

- Imprecise interventions: Imprecisely intervening means that the experimenter has some doubts concerning the target values of his manipulation unlike the case of standard interventions where he totally controls his manipulation.

  The probabilistic framework cannot solve such cases. Fortunately, the belief function framework is an appropriate tool to model such manipulations having imprecise target values. Besides, it allows to express beliefs in terms of subsets instead of singletons.

  Standard interventions are a particular case of imprecise interventions when the the subset representing the possible target values is composed of one element, i.e. $sub_{ij} = \{a_{ij}\}$.

- Uncertain interventions: An intervention may uncertainly happened by forcing A_i to take a specific unknown value a_{ij} ($a_{ij} \in \Theta_{A_i}$) or does not take place.

The "do" operator to represent interventions

The use of conditioning is appropriate when an event occurs spontaneously but inappropriate to compute the effects of external actions. The "do" operator allows to assign causal relations by differentiating the effects of observations and the effects of interventions.

In fact, interventional data allow the reasoning in a causal way by the mean of the "do" operator, originally introduced by (Goldszmidt & Pearl, 1992) for the ordinal conditional functions of (Spohn, 1988). This operator is proposed after that in (Pearl, 2000), following previous work by (Spirtes, 2001) under a probabilistic framework as a tool to represent interventions on an event that lets the manipulated event independent of all its initial causes. Then, (Boukhris, Elouedi, & Benferhat, 2011b) propose a counterpart of this operator to handle intervention under a belief framework.

An intervention on a variable A_i forcing it to take the value a_i is denoted $do(A_i = a_i)$ or $do(a_i)$. Our beliefs over the direct causes of a_i will not be modified. While conditioning is used to compute the effect of observations, the "do" operator is used as a tool to compute those of interventions on causal networks.

2.4 Causal Bayesian networks

The causal bayesian network becomes popular as an analytical framework in causal studies, where causal relations are encoded by the structure of the network. Several researchers have proposed a causal interpretation for Bayesian networks (Verma & Pearl, 1990), (Spirtes, 2001). Thus, a Bayesian network is used to model a probability distribution dictated by observations while a causal Bayesian network allows to model a probability distribution dictated by observations and interventions.

All probabilistic models describe a distribution over possible observed events, but they say nothing about what will happen if an intervention occurs. A causal network is a Bayesian network with the added property that the parents of each node are its direct causes and edges represent causal relationships.

A Bayesian network is a formalism that allows to represent compactly probability of a joint distribution over a set of variables using the concept of independence. However, a probability distribution, can be represented by several equivalent Bayesian networks if they describe exactly the same conditional independence relation and induce the same joint distributions due to Markov equivalence. Only one of these networks follows the causal process (see Figure 4.6), the so-called causal network.

Example 2.4 *Let us consider the networks presented in Figure 2.1 and continue with the same facts mentioned in Example 2.3. These networks are the same conditional independence relation but only one of them represent a causal process.*

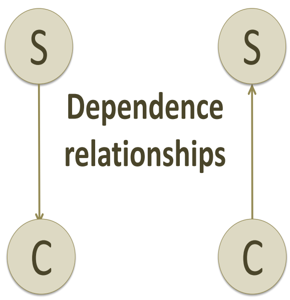

Figure 2.1: Associational equivalent networks

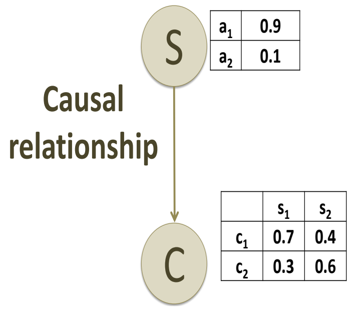

Figure 2.2: A causal Bayesian network

A causal Bayesian network (Pearl, 2000) can be defined in two components as follows:

- A graphical component: represented by a DAG, G = (V;E) where V is the set of variables and E is the set of edges encoding the dependencies among variables and describe cause-effect relations where an event is a cause of its child node and an effect of its parent node. The set of parent of A_i is denoted by $Pa(A_i)$. Arrows indicate only that one variable is causally relevant to another, and say nothing about the way in which it is relevant.

- A numerical component: representing conditional probability associated to each node(i.e., random variable) that quantifies the effects of its parent on it. Local probability distributions (a priori) defined for each node A_i in the context of its parent ($PA(A_i)$) must satisfy the following normalization constraints:
  - if A_i is a root node, then the a priori probability distribution of A_i should satisfy:

$$\sum_{a_i} P(a_i) = 1, \quad where \quad a_i \in \Theta_{A_i} \tag{2.1}$$

  - if $PA(A_i) \neq \emptyset$, then the conditional probability over A_i is defined as:

$$\sum_{a_i} P(a_i|Pa(A_i)) = 1, \quad where \quad Pa(A_i) \in \times\Theta_{A_j, A_j \in PA(A_i)} \tag{2.2}$$

On a Bayesian causal network, we can model the effects of not only observations but also those of interventions using the "do" operator (see section 2.2). The effect of an intervention $do(a_i)$ on the joint distribution is computed as follows:

$$P(a_1, \ldots, a_n|do(a_i)) = \frac{P(a_1, \ldots, a_n)}{P(a_i|Pa(A_i))} = P(a_1, \ldots, a_n|a_i, Pa(A_i)) \times P(Pa(A_i)$$

Example 2.5 *Let us continue with the causal network presented in Figure 4.6. Let us compare between the effect of seeing that someone has lung cancer (i.e., the variable C takes spontaneously the value c_1) and the effect upon acting on the variable C, by forcing the person to have lung cancer (forcing C to take the specific value c_1).*

- The effect of observing C with the value c_1 is computed with:

$$P(s_1, c_1|see(c_1)) = P(s_1, c_1|c_1) = \frac{P(s_1, c_1)}{P(c_1)} = \frac{0.63}{0.67} = 0.94$$

- The effect of acting on T forcing it to take the value c_1 is computed with:

$$P(s_1, c_1|do(c_1)) = \frac{P(s_1, c_1)}{c_1|Pa(c_1)} = \frac{P(s_1, c_1)}{P(c_1|s_1)} = P(s_1) = 0.9$$

Graphical representation of interventions

After acting on a variable, we assume that its initial causes are no more responsible of its state. Accordingly, an external action is handled.
There are mainly two methods for handling interventions by making changes on the structure of networks: graph mutilation and graph augmentation. These approaches have been proved to be equivalent under the probabilistic framework (Pearl, 2000), the possibilistic framework (Benferhat & Smaoui, 2007) and the belief framework (Boukhris, Elouedi, & Benferhat, 2011b).

Graph mutilation

An external action will alter the system. Thus, an intervention is interpreted by cutting off the edges pointed to the node concerned by the action. The rest of the network remains unchanged. The resulting graph is called **a mutilated graph denoted** by G_{mut}.

This action makes the direct causes (parent) of the variable concerned by the intervention no more responsible of its state. However, beliefs on its direct causes should not be modified.

Let G=(V,E) be a causal network and let A_i be a variable in G forced to take the value a_i by the intervention denoted $do(a_i)$. We define mutilation on two steps:

- Links (arcs) between A_i and their parents U_i will be deleted and its associated distribution is denoted $P_{G_{mut}}$. The resulting graph is denoted G_{mut}. The effect of this intervention on the joint distribution is represented by the new joint $P_{G_{mut}}(.|do(a_i))$.
  In the mutilated graph, it corresponds to observing $A_i = a_i$. Thus, it simply consists of conditioning the mutilated graph by the value a_i.

$$P_{G_{mut}}(.|a_i) = P_G(.|do(a_i)) \tag{2.3}$$

- An action $do(a_i)$ impose the value a_i on a variable A_i. The corresponding distribution of A_i is defined as follows:

$$P(a_k) = \begin{cases} 1 & if \quad a_k = a_i \\ 0 & otherwise \end{cases} \tag{2.4}$$

Example 2.6 *Let us consider by S the fact of smoking,* $\Theta_S = \{s_1, s_2\}$ *where* s_1 *is yes and* s_2 *is no and by C the fact of having lung cancer,* $\Theta_C = \{c_1, c_2\}$ *where* c_1 *is yes and* c_2 *is no. These two events represent a description of the relation between smoking and having lung cancer. After acting on a variable by forcing C to take the value* c_1*, the state of C will be independent from the fact of smoking (S). Therefore, the link relating S to C will be deleted. This is represented by the graph in Figure 2.3*

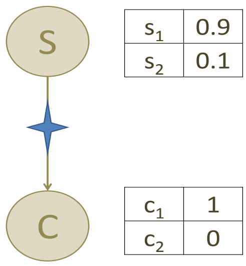

Figure 2.3: Graph mutilation upon the intervention $do(c_1)$

graph augmentation

An alternative but equivalent approach in order to represent interventions. The idea consists of considering an intervention as an extra node in the system by adding this fictive node called "DO" to the variable A_i concerned by an intervention. The parents set of the variable A_i denoted PA is augmented by the extra node $do(a_i)$ and becomes PA'=PA$\cup$ DO. The resulting graph is called **an augmented graph** and denoted by G_{aug}. This method allows to represent the effect of observations and interventions.

The DO node is taking value in do(x), x$\in \{\Theta_{A_i} \cup \{\text{nothing}\}\}$. Do(nothing) means that there are no actions on the variable A_i, it represents the state of the system when no interventions are made. $Do(a_i)$ means that the variable A_i is forced to take the value a_i. The distribution of this node is defined by:

$$P(do(x)) = \begin{cases} 1 & if \quad x = nothing \\ 0 & otherwise \end{cases} \tag{2.5}$$

Values $do(a_i)$ mean that the value A_i is forced to take the value a_i. The distribution relative to the node DO is defined as follows:

$$P(do(x)) = \begin{cases} 1 & if \quad x = a_i \\ 0 & otherwise \end{cases} \tag{2.6}$$

The new local distribution relative to A_i after augmenting the graph, resulting from adding the extra node DO which is a parent of A_i, is given by:

$$P(a_k|Pa(A_i), do(x)) = \begin{cases} 1 \quad if \quad x = a_i \\ 0 \quad if \quad x \neq a_i \\ P(a_k|Pa(A_i)) \quad if \quad x = nothing \end{cases} \tag{2.7}$$

Example 2.5 *Let consider the same facts, smoking (S) and lung Cancer (C) mentioned in Example 2.3. After acting on a variable by forcing C to take the value c_1. Therefore, the set of parents of C becomes $S \cup DO$ where the node DO is set to the value $do(c_1)$. This is represented by the graph in Figure 2.4.*

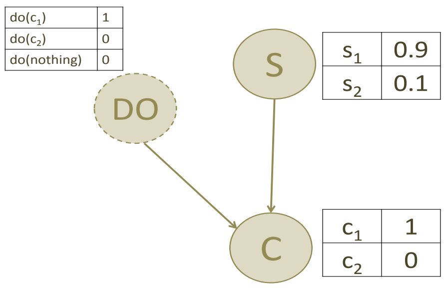

Figure 2.4: Graph augmentation upon the intervention $do(c_1)$

2.5 Causal possibilistic network

An alternative causal model under a possibilistic framework was also proposed showing its efficiency especially when cases require pure qualitative and ordinal handling (Benferhat & Smaoui, 2007).

Possibilistic networks represent an effective tool for the representation and processing of uncertain, incomplete and imprecise information. As in the case for Bayesian networks, the identification of causal relationships and the management of external interventions in possibilistic networks system requires additional properties. As for probabilistic networks, a joint possibility distribution can be represented by several possibilistic networks equivalent in terms of independence relationships, but one can follow the direction of the causal process. Such a network is called causal possibilistic network.

Causal possibilistic networks (Benferhat, 2010), (Benferhat & Smaoui, 2011) were developed to model causal knowledge under the possibilistic framework where directed arcs of the graph are interpreted as representing causal relations between events. Arcs also follow the direction of causal process. Intuitively, the parent set Pa_i of A_i represents all the direct causes for A_i.

Causal possibilistic networks represent an efficient way to deal with uncertain data and to predict evolution of the system under effects of interventions which are computed using an adaptation of the "do" operator to a possibilistic framework(Benferhat & Smaoui, 2007). As for causal Bayesian networks, interventions can be handled either by means of mutilated graphs or by means of augmented possibilistic causal networks. These two approaches are equivalent to graphically represent such manipulations under the possibilistic network in possibilistic causal networks.

2.6 Causal belief networks

Causal belief networks (Boukhris, Elouedi, & Benferhat, 2011b), based on the belief network (BNC) presented in Chapter 1, is a model under an uncertain environment where the uncertainty is represented by belief masses. This graphical model represent an alternative and an extension to Bayesian causal networks, that offer interesting tools to handle interventions. It allows the detection of causal relationships under the belief function framework resulting from acting on some events.

In order to predict the effects of external actions on the system, the construction of the belief causal network must be different from belief network and the conditioning on observation should be distinguished from a conditioning on an external action. Handling interventions and computing their effects on the system can be done by making changes on the structure of the belief causal network. The two equivalent methods developed were namely, belief graph mutilation and beleif graph augmentation methods (Boukhris, Elouedi, & Benferhat, 2011b).

2.7 Conclusion

In this chapter, we have presented an overview of some definitions of causality. Since an intervention is a crucial concept for an efficient causal analysis, we have exposed their types and then tools to model intervention on causal belief networks and provide a survey of its graphical model representations. We also detailed the existing belief causal networks which are appropriate tools to model causal knowledge which is usually uncertain. We explained their usefulness to represent interventions and compute their effects.

Existing algorithms deal with the propagation of observational data in belief networks. Unfortunately, there is not any algorithm for belief causal inference, i.e., dealing with effects of observational and interventional data. Therefore, in the next chapter, we will investigate causal inference in singly connected graphs under a belief function framework and we will propose new algorithms to propagate information.

Chapter 3

Inference in singly connected causal belief networks

3.1 Introduction

The impact of a new information on the remaining variables can found by first computing the joint distribution and then making marginalization. This method is not suitable when the number of variables becomes important. To solve this problem, equivalent local computations have been proposed (Pearl, 1988), (Shachter, 1988), (Lauritzen & Shenoy, 1996).

Some works have investigated causal inference (Pearl, 2000), (Benferhat & Smaoui, 2011) respectively within probability theory and possibility theory. For belief networks, existing algorithms only deal with the propagation of observational data in belief networks (Ben Yaghlane & Mellouli, 2008), (Xu & Smets, 1996). This can be done by using two rules proposed by Smets (1993b) namely, disjunctive rule of combination (DRC) and generalized Bayesian theorem (GBT). Unfortunately, there is not any algorithm for belief causal inference. Since, in (Boukhris, Elouedi, & Benferhat, 2011b), it was shown that interventions can be graphically handled using the belief graph mutilation and the belief graph augmentation methods, we have proposed new algorithms able to handle not only observational but also interventional data in singly connected networks where belief are defined per edge: an algorithm adapted from (Ben Yaghlane & Mellouli, 2008) to propagate an observation and an intervention in causal networks and another extended from standard Pearl's algorithm (Pearl, 1988) which is more suitable in the case where we have several observations and interventions. Our proposed algorithms are based on the two rules proposed by Smets. The basic operation during the inference process is the combination.

In this chapter, we recall in Section 3.2 Pearl's (1988) probabilistic propagation algorithm so-called the standard propagation and the extension of this algorithm proposed by (Peot & Shachter, 1991) so-called the centralized propagation. Then, in Section 3.3, we describe the basic algorithm proposed by (Ben Yaghlane & Mellouli, 2008) to handle observational data which an adaptation of Pearl's centralized algorithm. In Section 3.4, we propose a centralized propagation algorithm in causal belief networks to handle observations and also interventions based on the mutilation and the augmentation of the graph. Finally, in Section 3.5, we propose another manner for causal inference that is extended from standard Pearl's algorithm.
Note that all following algorithms are defined for the case of singly connected networks (also known as polytree) where there is at most one directed path between any two nodes in the graph.

3.2 Pearl's probabilistic propagation algorithm

There are several algorithms dealing with propagation in Bayesian networks. Some algorithms are proposed to propagate information in singly connected bayesian networks (Pearl, 1988), (Kim & Pearl, 1983) while others to make inference in multiply connected bayesian networks (Lauritzen & Spiegelhalter, 1988), (Jensen, 1996).

Since we were focusing on singly connected networks, this section is dedicated to give an overview of Pearl's standard algorithm (Pearl, 1988) and Pearl's centralized algorithm (Peot & Shachter, 1991). The main principle underlying these algorithms is the use of probabilistic independence properties coded in the structure of networks. They consist in finding the impact of a new information (evidence e) on the remaining variables. This evidence (observation) perturbs the system and has to be propagated through the network via message-passing between neighboring variables. Note that conditional beliefs are defined for all parents.

Some elements of the message-passing scheme that have been used in this part are recalled here:

- U the set of nodes
- Pa(X) $\subseteq$ U is the set of all X's parents
- Ch(X) $\subseteq$ U is the set of all X's children
- When receiving a message each node X updates both local vectors:
  - $\pi(x_1, ..., x_n)$: the vector concerning messages received by its parents.

  - $\lambda(x_1, ..., x_n)$: the vector concerning messages received by its children.

- Each node sends and receives messages from each of its neighbors. The local message-passing between variables is based on two kinds of messages:
  - π-message: a message sent from a parent node to a child node.
  - λ-message: a message sent from a child node to a parent node.
- e_X^+ presents the set of all causes of X (his parents and his non-descendants).
- e_X^- presents the set of all effects of X (X and his descendants).

These notations lead to the following recursive expressions for computing values and messages:

$$BEL(x) = P(x|e) = \alpha.\pi(x).\lambda(x) \tag{3.1}$$

$$\lambda(x) = P(x|e_X^-) = \lambda_X(x).\prod_{j=1..m} \lambda_{Y_j}(x) \tag{3.2}$$

$$\pi(x) = P(x|e_X^+) = \sum_u P(x|u).\prod_{i=1..n} \pi_X(U_i) \tag{3.3}$$

$$\lambda_X(U_i) = \sum_x \lambda(x)[\sum_{U_K:K\neq i} P(x|u).\prod_{K\neq i} \pi_X(U_K)] \tag{3.4}$$

$$\pi_{Y_j}(x) = \lambda_X(x).\prod_{i=1..m,i\neq j} \lambda_{Y_i}(x).\pi(x) \tag{3.5}$$

where:

- BEL(x) is the current conditional probability based on the total evidence e.
- α is the factor of normalization.
- $\pi(x)$ is the π value $\forall x \in \Theta_X$.
- $\lambda(x)$ is the λ value $\forall x \in \Theta_X$.
- $\lambda_X(U_i)$ is the message from X to its parent U_i where $U_i = u_i$.
- $\pi_{Y_j}(x)$ is the message from X to its child Y_j where X=x.

3.2.1 Standard Pearl's algorithm

This improved version decreases the number of messages since each node is visited at most twice regardless of the number of observed nodes.

To make inference using the standard Pearl's algorithm (Pearl, 1988), these steps should followed:

Step1: Initialization
In this step, the algorithm initializes all fields π using equation(3.2), *lambda* using equation(3.3) and BEL using equation(3.1).

Step2: Propagation-up
Algorithm applies propagation in direction of leaves node until reaching roots using equation(3.4) where each node sends messages to its parents which in turn send messages to their parents.

Step3: Propagation-down
Algorithm applies propagation in direction of roots until reaching leaves using equation(3.5) where each node send messages to its children which in turn send messages to their children.

3.2.2 Centralized Pearl's algorithm

(Peot & Shachter, 1991) proposed a centralized version of the standard algorithm of Pearl. The principle of this version is to choose a pivot node to determine the direction in which messages should be sent. The main steps of this algorithm are as follows:

- Initialization: All vectors (π, λ) and the local evidence (λ_X) are initialized.
- Collect-evidence: Each node sends a message to each of its neighbors in the direction of the pivot.
- Distribute-evidence: The pivot sends a message to its neighbors which in turn sends a message to its neighbors.
- Marginalization: Each node computes its BEL.

The algorithm of propagation is as follows:

<u>Initialization</u>
$\lambda \leftarrow 1$
$\pi \leftarrow 1$
$\lambda_X \leftarrow 1$
For each root X : set $\pi_X \leftarrow P_A$
For each observed node X :
if A is instantiated to a, $\lambda_X \leftarrow 1$, $\pi_X \leftarrow 0$.
<u>Collect-evidence</u>
For i from 1 to length(Postordre) - 1
  X $\leftarrow$ Post-order[i]
  Y $\leftarrow$ adjacent node to X in Postorder
  compute $\lambda(X)$ using (3.2), calculate π using equation(3.3).
  if Y is a parent of X then send a message λ from A to Y using equation(3.4).
  else send a un message π from A to B using equation(3.5).
End for
<u>Distribute-evidence</u>
For i from 1 to length(Pre-order) - 1
  X $\leftarrow$ Pre-order[i]
  Y $\leftarrow$ adjacent node to X in Pre-order
  compute $\lambda(X)$ using equation(3.2), calculate π using equation(3.3).
  if Y is a parent of X then send a message λ from A to Y using equation(3.4).
  else send a un message π from A to B using equation(3.5).
End for
<u>Marginalization</u>
For each node, we calculate BEL(A)=$P(a|e)$ using equation(3.1).

3.3 Propagation in belief networks

Several algorithms for inference in belief networks are proposed by Xu et al. (1996), Ben Yaghlane et al. (2008) which is an extension of Xu et al. algorithm and Simon et al. (2008). Since, we deal with singly connected networks, we will present in this section the algorithm of (Ben Yaghlane & Mellouli, 2008).

The belief propagation algorithm (Ben Yaghlane & Mellouli, 2008) is inspired from Pearls centralized algorithm. In order to propagate information via the network, (Ben Yaghlane et al., 1998) uses the DRC (Disjunctive Rule of Combination) and GBT (Generalized Bayesian Theorem) to make possible the knowledge representation by conditional belief functions and to use them directly for reasoning in the DEVN. Note that conditional beliefs are defined per single parent.

3.3.1 Disjunctive rule of combination (DRC)

Let $\{pl^X(.|y_i) : y_i \in \Theta_Y\}$ be the set of conditional plausibilities given Θ_Y. Suppose m_0^Y is the bba collected at node Y that must be propagated in order to compute pl_1^X.
Using the GBT, we have:

$$pl^X(x) = \sum_{y \subseteq Y} m_0^Y(y)(1 - \prod_{y_i \in y}(1 - pl^X(x|y_i))) \tag{3.6}$$

Example 3.1 *Let $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$. The initial vectors are m_Y and $pl^X(.|y_i) : i = 1, 2, 3$. The column $pl^X(.|y)$ is computed using equation(3.6). Table 3.1 presents an example of DRC propagation.*

Table 3.1: DRC propagation

	$\{y_1\}$	$\{y_2\}$	$\{y_3\}$	$\{y_1,y_2\}$	$\{y_1,y_3\}$	$\{y_2,y_3\}$	Θ_Y	$pl^X(.\|y)$
ø	0	0	0	0	0	0	0	0
$\{x_1\}$	1	0.2	0.85	1	1	0.82	1	0.14
$\{x_2\}$	0.7	1	1	0.3	1	1	1	0.32
Θ_Y	1	1	1	1	1	1	1	0.54
m^Y	0.2	0.4	0	0	0.3	0	0.1	

3.3.2 Generalized Bayesian Theorem (GBT)

Let $\{pl^X(.|y_i) : y_i \in \Theta_Y\}$ be the set of conditional plausibilities given Θ_Y. Suppose m_0^X is the bba collected at node X that must be propagated in order to compute pl_1^Y.
Using the DRC, we have:

$$pl^Y(y) = \sum_{x \subseteq X} m_0^X(x)(1 - \prod_{y_i \in y}(1 - pl^X(x|y_i))) \tag{3.7}$$

Example 3.2 *Let $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$. m_X and $pl^X(.|y_i) : i = 1, 2, 3$ are initial vectors. The line $pl^Y(.|x)$ is computed using equation (3.7). Table 3.2 presents an example of GBT propagation.*

Table 3.2: GBT propagation

	m^X	$\{y_1\}$	$\{y_2\}$	$\{y_3\}$	$\{y_1,y_2\}$	$\{y_1,y_3\}$	$\{y_2,y_3\}$	Θ_Y
$\emptyset$	0	0	0	0	0	0	0	0
$\{x_1\}$	0.2	1	0.2	0.85	1	1	0.82	1
$\{x_2\}$	0.7	0.7	1	1	0.3	1	1	1
Θ_X	0.1	1	1	1	1	1	1	1
$pl^Y(.\|x)$		0.02	0.18	0.4	0	0.14	0.04	0.22

3.3.3 Propagation algorithm in belief networks

Let U be a finite set of nodes of a given network, r be a the set of roots. For each X $\in$ U, let Pa(X) $\subseteq$ U be the set of parents of X, and Ch(X) $\subseteq$ U be the set of children of X. For each node X, we have to store its a priori belief function bel_0^X. Also, we store conditional belief functions for each node given its parents $\{bel^X(.|Y) : Y \in Pa(X)\}$.

The algorithm proposed by (Ben Yaghlane & Mellouli, 2008) consists in an initialization step and on updating step such that:

- **<u>Initialization</u>**

  In this step, messages propagate from roots until reaching leaves. When a node X receives a message from its parents, it will be able to compute and send a message π to its children Y using the following formula:

  $$\pi_Y = bel^Y \oplus (\oplus_{X \in Pa(Y)} \pi_{X \to Y}) \tag{3.8}$$

  Then, Y computes the its a priori belief function using this formula:

  $$bel^Y \leftarrow \pi_Y \oplus \lambda_Y \tag{3.9}$$

Algorithm.Initialization
For i from 1 to length(U)
  set $bel^X \leftarrow m_0^X$
  set $\pi_X \leftarrow m^X$
  set $\lambda_X \leftarrow$ vacuous belief
End for
For i from 1 to length(r)
  send $\pi_{X \to Y}$ message for all children Y of X using DRC equation(3.6).
End for
For i from 1 to Ch(X)
  Node Y has to wait until it holds the messages of all its parents
  Compute the new π value using (equation(3.8))
  Compute the new marginal using (equation(3.9))
  Send a new π_Y message for all its children using DRC equation(3.6).
End for

- **Updating**

When a new observation is introduced into a given node, the updating algorithm will be performed. We assume that if a node X is instantiated then we have a new observation (O_Y), otherwise O_Y is a vacuous belief function. The algorithm consists in visiting all nodes. When we visit a node X, we compute all the incoming messages. Then, the node X will be able to compute its π_Y value, its λ_Y value, its new marginal bel^Y. Thus, it can compute and send all outgoing messages.

  - **Computing values**

    * π_Y value is obtained by the initial marginal value with the messages coming from all its parents.

$$\pi_Y = bel_0^Y \oplus (\oplus_{X \in Pa(Y)} \pi_{X \to Y}) \tag{3.10}$$

    * λ_Y value is obtained by the new observation (O_Y) with the messages coming from all its children.

$$\lambda_Y = O_Y \oplus (\oplus_{Z \in Ch(Y)} \lambda_{Z \to X}) \tag{3.11}$$

    * The new marginal m_X is obtained by combining the π_X value and the λ_X value.

$$bel^Y = \pi_Y \oplus \lambda_Y \tag{3.12}$$

– **Computing messages**

Once the node Y is updated, then it will send new messages to all its neighbors that not have been updated.

* $\pi_{Y \to Z}$ representing the message sent from a node Y to its children Z using DRC (equation(3.6)).
* $\lambda_{Y \to X}$ representing the message sent from a node Y to its parents using GBT (equation(3.7)).

Algorithm. Updating
```
Repeat
  Y computes its new value bel^Y using equation(3.10).
  For i from 1 to length(Ch(Y))
    compute and send a new message π_{Y→Ch(Y)(i)} using DRC equation(3.6)
    to all children yet been updated.
  End for
  For i from 1 to length(Pa(Y))
    compute and send a new message λ_{Y→Pa(Y)(i)} using GBT equation(3.7)
    to all children not yet been updated.
  End for
Until there is no nodes to update.
```

3.4 Centralized propagation in causal belief networks

In this section, we propose new propagation algorithms for singly connected causal belief networks where beliefs are quantified by conditional masses and defined per edge. Our proposed algorithms allow to evaluate the effect of interventions and observations with different ways, i.e., mutilating the graph or augmenting it.

3.4.1 Propagating observations in causal belief networks.

In this section, we have proposed an adaptation of the algorithm proposed by (Ben Yaghlane & Mellouli, 2008) to propagate observations in causal belief networks.

Initialization

In our proposed algorithm, each node sends messages to its children. Each child sends in turn a message to its children. Since a child has to wait until it holds the messages of all its parents, to move from one level to another, we have to make sure that all nodes are updated. For that, we propose that all nodes send messages in the direction of the leaves after the receiving all messages from their parents. This direction is called post-order. Each node updates its values m^X, π_X and λ_X and each child updates its π value using this formula:

$$\pi_Y = m^Y \oplus (\oplus_{X \in Pa(Y)} \pi_{X \to Y}) \tag{3.13}$$

Then, this node has to compute its marginal by using this formula:

$$m^Y \leftarrow \pi_Y \oplus \lambda_Y \tag{3.14}$$

Algorithm. Initialization
For i from 1 to length(Post-order)
  set $m^X \leftarrow m_0^X$
  set $\pi_X \leftarrow m^X$
  set $\lambda_X \leftarrow$ vacuous bba
  send $\pi_{X \to Y}$ message for all children Y of X using equation(3.6).
  for i from 1 to length(Ch(X))
    Y has to wait until it holds the messages of all its parents then,
    1) compute the new π value using equation(3.13).
    2) compute the new marginal using equation(3.14).
    3) send a new π_Y message for all its children using equation(3.6).
  End for
End for

Example 3.3 *Let us consider the following directed causal belief network in Figure 3.1 constituted by 6 nodes U={A,B,C,D,E,F} representing the variables of the problem. For the sake of computational simplicity, all the variables used in this example are binary.*

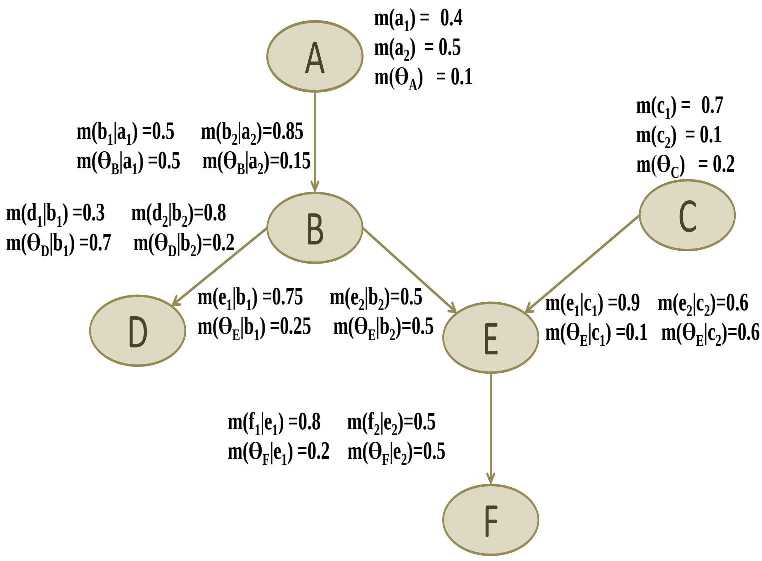

Figure 3.1: Initialization

Before that any variable will be instantiated, we perform the initialization step in which we apply the propagation method described above. The order of propagation is $\{A, B, C, E, D, F\}$. According to this order, we will choose to initialize A B, C, E, D and finally F.

Node A

- *Compute and send $\pi_{A\to B}$ message to B using equation(3.6). So, for any $b \in \Theta_B$, $m_{A\to B}(b)$ is given as follows:*
  $m_{A\to B}(b_1) = 0.2 \quad m_{A\to B}(b_2) = 0.425 \quad m_{A\to B}(\Theta_B) = 0.375.$

Node B *When B receives a new message from its parent A:*

- *Compute its π value using equation(3.13) and its new marginal using equation(3.14) in this manner: $\pi_B = m(B) = \pi_{A\to B}$*
  $m_{A\to B}(b_1) = 0.2 \quad m_{A\to B}(b_2) = 0.425 \quad m_{A\to B}(\Theta_B) = 0.375.$
- *Compute and send $\pi_{B\to D}$ using B message to D using equation(3.6). So, $m_{B\to D}$ is as follows:*
  $m_{B\to D}(d_1) = 0.06 \quad m_{B\to D}(d_2) = 0.34 \quad m_{B\to D}(\Theta_D) = 0.6.$
- *Similarly, compute and send $\pi_{B\to E}$ message to E using equation(3.6). So, $m_{B\to E}$ is as follows:*
  $m_{B\to E}(e_1) = 0.15 \quad m_{B\to E}(e_2) = 0.21 \quad m_{B\to E}(\Theta_E) = 0.64.$

Node C

- *C sends a new message* π *to E using equation(3.6). Then, E updates its values using equation(3.13). and equation (3.14) respectively.*
  $m_{C\to E}(e_1) = 0.63 \quad m_{C\to E}(e_2) = 0.06 \quad m_{C\to E}(\Theta_E) = 0.31.$

Node E *E receives new messages from its parents B and C.*

- *Compute its* π *value using equation(3.13) and its new marginal using equation(3.14).*
  First we combine all π *messages coming from parents using* $\pi_E = \pi_{B\to E} \oplus \pi_{C\to E}$. *We obtain* $\pi_E(e_1) = 0.633 \quad \pi_E(e_2) = 0.136 \quad \pi_E(\Theta_E) = 0.231.$

- *Then, we combine* π *with its bba and its* λ *using equation(3.14) and we obtain:*
  $m_E(e_1) = 0.633 \quad m_E(e_2) = 0.136 \quad m_E(\Theta_E) = 0.231.$

- *Compute and send* $\pi_{E\to F}$ *message to F using equation(3.6).*

Node F *When F receives new messages from its parent E.*

- *Compute its* π *using equation(3.13) and its new marginal using equation(3.14).*
  $m_F(f_1) = 0.507 \quad m_F(f_2) = 0.068 \quad m_F(\Theta_F) = 0.425.$

Finally, the results of initialization are illustrated in Figure 3.2:

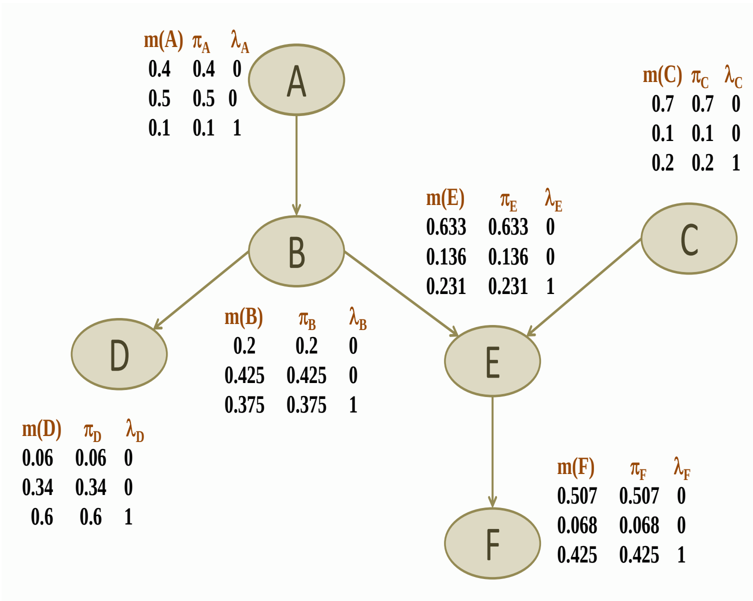

Figure 3.2: The state of the causal belief network after initialization

Updating

This process starts when new observation is performed at a given node. The algorithm is based on computing new marginal m_Y by combining the new message π if the the receiving is a parent or a message λ if the node is a child with the previous marginal which combine all messages coming from neighbors and the initial a priori.

When a new observation (O_Y) is introduced, the node Y concerned by the intervention updates its new marginal by combining its previous marginal m_p^Y with the new observation (O_Y).

$$m^Y = O_Y \oplus m_p^Y \tag{3.15}$$

Message passing is depending on the direction in which the message is circulated. Initially, messages propagate from the node concerned by the observation to other nodes. Any node which has received messages from its neighbor sends an other message to the its remaining neighbor not yet been updated and updates its values π or λ and its new marginal.

$$m^Y = \pi^Y \oplus m_p^Y \quad \text{If the message is coming from a parent} \tag{3.16}$$

$$m^Y = \lambda^Y \oplus m_p^Y \quad \text{If the message is coming from a child} \tag{3.17}$$

As for the initialization process, to ensure the updating process, we have to define an order A of variables to perform message-passing between neighbors.

Once the node Y is updated, then it will send new messages to all its neighbors.

- $\pi_{Y \to Z}$ representing the message sent from a node Y to its children Z using the equation(3.6).

- $\lambda_{Y \to X}$ representing the message sent from a node Y to its parents using equation(3.7).

Here is the algorithm corresponding to the updating process, it consists on a recursive procedure which propagate information through the totality of the network:

Algorithm. Updating
Y computes its new marginal using equation (3.15).
Y sends messages to all its neighbors using equation (3.6) or equation (3.7).
Define an order of propagation A.
For i from 1 to length(A)
  A(i) computes its new marginal m^Y using equation (3.16) or equation (3.17).
  For i from 1 to length (ne (A(i)))
    Send a new message using equation(3.6) or equation (3.7) to all neighbors not yet been updated.
  End for
End for

Example 3.3 (continued)
Suppose now that new observation (O_B) is introduced at a node B, where $O_B = (1, 0, 0)$. So, node B is instantiated. For updating the causal belief network, the propagation algorithm described above will be performed.

Node B *If B is instantiated, then an order should be defined for the propagation of this evidence. The proposed order is {B, A, D, E, C, F}*

- *Compute the new values for the node B.*
  So, $m(b_1) = 1 \quad m(b_2) = 0 \quad m(\Theta_B) = 0$
- *Sent a new message to all its neighbors (A, D and E):* $\pi_{B\to D}$ *message to D,* $\pi_{B\to E}$ *message to E and* $\lambda_{B\to A}$ *message to A.*
  $\lambda_{B\to A}(a_1) = 0.85 \quad \pi_{B\to D}(d_1) = 0.3 \quad \pi_{B\to E}(e_1) = 0.75$
  $\lambda_{B\to A}(a_2) = 0 \quad \pi_{B\to D}(d_1) = 0 \quad \pi_{B\to E}(e_2) = 0$
  $\lambda_{B\to A}(\Theta_A) = 0.15 \quad \pi_{B\to D}(\Theta_D) = 0.7 \quad \pi_{B\to E}(\Theta_E) = 0.25$

Node A

- *Node A computes its new marginal by combining its previous marginal with this new message π using equation 3.15*
  where $m(a_1) = 0.84 \quad m(a_2) = 0.13 \quad m(\Theta_A) = 0.03$
- *A has no children and no parents not yet been updated.*

Node D

- *Node D computes its new marginal by combining its previous marginal with this new message π.*
  where $m(d_1) = 0.27 \quad m(d_2) = 0.26 \quad m(\Theta_D) = 0.47$

- *D has no children and no parents not yet been updated.*

Node E

- *Node E computes its new marginal and the result is as follows:*
  where $m(E_1) = 0.27 \quad m(E_2) = 0.26 \quad m(\Theta_E) = 0.47$

- *D has a child F and a parent C. So, it will send messages to its neighbor.*

Node C

- *Node E computes its new value* λ *where*
  $\lambda_{E \to C}(c_1) = 0.54 \quad \pi_{E \to C}(c_2) = 0.34 \quad \pi_{E \to C}(\Theta_C) = 0.43$
  Then, it computes its new marginal and the result is as follows:
  where $m(E_1) = 0.85 \quad m(E_2) = 0.05 \quad m(\Theta_E) = 0.1$

- *C has no children and no parents not yet been updated.*

Node F

- *Node E computes its new value* λ *where*
  $\lambda_{E \to C}(c_1) = 0.54 \quad \pi_{E \to C}(c_2) = 0.34 \quad \pi_{E \to C}(\Theta_C) = 0.43$
  Then, it computes its new marginal and the result is as follows:
  where $m(E_1) = 0.05 \quad m(E_2) = 0.07 \quad m(\Theta_E) = 0.43$

- *C has no children and no parents not yet been updated.*

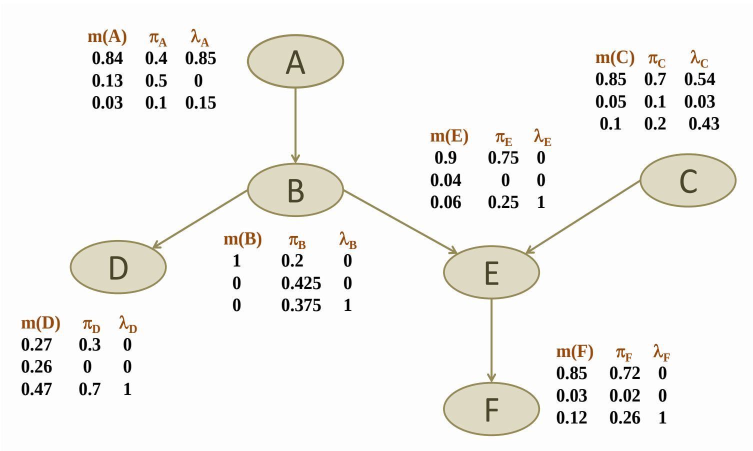

Figure 3.3: The causal belief network upon the observation b_1 observation.

3.4.2 Propagation in the mutilated graph

One possible way to handle interventions is to mutilate the graph. To perform causal inference and compute the effect of an intervention on the mutilated graph, we could simply make a conditioning on the mutilated graph by the target value on the variable concerned by the intervention.

As for the case of the propagation of evidence described above, the algorithm of propagation in the mutilated graph is based on two steps: initialization and propagation.

Initialization process

The initialization process is the same for the case of observation (see section 3.1)

Updating process

When new intervention I is introduced at a given node Y, the updating algorithm will be performed. All edges pointing the the node Y will be deleted. Then Y becomes a root. It computes its new marginal by combining its previous marginal m_p^Y with the new intervention (I_Y).

The new marginal of Y is computed as follows:

$$m^Y = (I_Y) \oplus m_p^Y \tag{3.18}$$

Once the node Y is updated, then it will send new messages to all its neighbors.

- $\pi_{Y \to Z}$ representing the message sent from a node Y to its children Z using the equation(3.6).

- $\lambda_{Y \to X}$ representing the message sent from a node Y to its parents using equation(3.7).

Algorithm. Updating
Cutting all edges pointing to Y.
Y becomes a root.
Y computes its new marginal using equation(3.18)
Define an order of propagation A.
For i from 1 to length(A)
  A(i) computes its new marginal m^Y using equation (3.16)
  or equation (3.17).
  For i from 1 to length (ne (A(i)))
    Send a new message using equation(3.6) or equation (3.7) to all
    neighbors not yet been updated.
  End for
End for

Example 3.3 (continued)
Suppose now that new intervention (I_B) is introduced at a node B, where $do(b_1) = 1$, $do(b_2) = 0$ and do(nothing)=0. Arc relating A to B will be delated. The propagation algorithm described above will be performed.

Node B *If B is instantiated, then an order will be defined for the propagation of the evidence. The proposed order is {B, A, D, E, C, F}*

- *Compute the new values for the node B.*
  So, $m(b_1) = 1 \quad m(b_2) = 0 \quad m(\Theta_B) = 0$

- *Send a new message to his neighbors (D and A): $\pi_{B\to D}$ message to D and $\pi_{B\to E}$ message to E*
  $\pi_{B\to D}(d_1) = 0.3 \quad \pi_{B\to D}(d_2) = 0 \quad \pi_{B\to D}(\Theta_D) = 0.7$ $\pi_{B\to E}(e_1) = 0.75 \quad \pi_{B\to E}(e_1) = 0 \quad \pi_{B\to E}(\Theta_E) = 0.25$

Node A *Beliefs on the node A will not be changed.*

Node D

- *Node D computes its new marginal by combining its previous marginal with this new message $\pi_{B\to D}$.*
  where $m(d_1) = 0.27 \quad m(d_2) = 0.26 \quad m(\Theta_D) = 0.47$

- *D has no children and no parents not yet been updated.*

Similarly, we do the same computations for node E, C and F respectively.

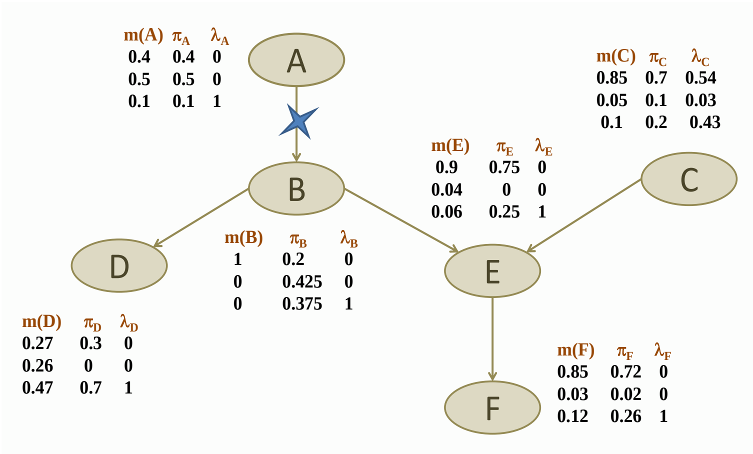

Figure 3.4: The causal belief network upon the intervention $do(b_1)$.

3.4.3 Propagation in the augmentation graph

An alternative way to handle interventions is to alter the structure of the graph by adding a new fictive node DO representing the intervention.

The DO node is taking value in $do(x), x \in \{\Theta_{A_i} \cup \{nothing\}$. do(nothing) represents the state of the system when no interventions are made. $do(a_{ij})$ means that the variable A_i is forced to take the value a_{ij}.

The main advantage of handling interventions by graph augmentation, is that it allows to represent the effect of interventions and also observations. To compute the effect of observations, we make a conditioning of the augmented graph by the value do(nothing) and to compute the effect of interventions, we make a conditioning of the augmented graph by the value $do(a_{ij})$.

As for standard causal belief networks, conditional distributions in the augmented belief networks are defined per single parent. Doing this way is assuming that conditional distribution can be defined by a different local source. Consequently, we will have a source given the DO node, i.e., the intervention and a source or multiple sources given the initial causes.
In the following, we will propose a definition of conditional distributions in the augmented belief graph when beliefs are defined per single parent.

Defining conditionals given the DO node

- In the case where no interventions are performed, the state of the variable concerned by the action is unknown. This situation represents the state of total ignorance. Therefore, the conditional distribution given the node "DO" becomes a vacuous bba. The new distribution of the node DO is defined as:

  – For each $sub_{ik} \subseteq \Theta_{A_i}$ and $x = \{nothing\}$

$$m^{A_i}_{G_{aug}}(sub_{ik}|do(nothing)) = \begin{cases} 1 & \text{if} \quad sub_{ik} = \Theta_A \\ 0 & \text{otherwise} \end{cases} \tag{3.19}$$

- In the case where there is an intervention that pushes a variable A_i to take a specific value, the conditional distribution given the node "DO" is defined as:

  – For each $sub_{ik} \subseteq \Theta_{A_i}$ and $x = \{a_{ij}\}$

$$m^{DO}_{G_{aug}}(sub_{ik}|do(a_{ij})) = \begin{cases} 1 & if \quad x = a_{ij} \\ 0 & otherwise \end{cases} \tag{3.20}$$

Defining conditionals given the initial causes

- In the case where there is no interventions, the initial parents of each node represent its direct causes. Therefore, the conditional beliefs of the augmented graph dealing with no interventions are the same for the initial graph where:

$$m^{A_i}_{G_{aug}}(sub_{ik}|Pa_j(A_i)) = m^{A_i}_{G}(sub_{ik}|Pa_j(A_i)) \tag{3.21}$$

- In the case where an intervention is performed: an intervention is an external action which completely controls the state of the target variable. So, the source who predicted that the distribution of the target variable is defined by the initial distribution is considered as totally unreliable ($\alpha = 1$) and should be weaken. In fact, the effect of the initial causes should no longer be considered.
  Hence, the new conditional bba of the target variable given the initial causes will be discounted:

$$m_{G_{aug}}(sub_{ik}|Pa(A_i)) = \begin{cases} \alpha & \text{if} \quad sub_{ik} = \Theta_A \\ 1-\alpha & \text{otherwise} \end{cases} \tag{3.22}$$

  Since the source is totally unreliable (α=1), the conditional bba becomes as follows:

$$m_{G_{aug}}(sub_{ik}|Pa(A_i)) = \begin{cases} 1 & \text{if} \quad sub_{ik} = \Theta_A \\ 0 & \text{otherwise} \end{cases} \tag{3.23}$$

Algorithm of propagation in the augmented graph

Initialization process

Initially, there is no interventions. So, we add the fictive node DO and set it to the value value do(nothing). Then, we defined the distribution of the "DO" node is defined using equation(2.3). Since there is no action that perturbs the system, the algorithm of initialization is similar to the one proposed to handle observation. However, we have to add the fictive node DO and assign it the value nothing.

Algorithm. Initialization
Add a fictive node which sends a message to its child.
For i from 1 to length(Post-order)
  set $m^X \leftarrow m_0^X$
  set $\pi_X \leftarrow m^X$
  set $\lambda_X \leftarrow$ vacuous bba
  send $\pi_{X \rightarrow Y}$ message for all children Y of X equation(3.6).
  for i from 1 to length(Ch(X))
    Y has to wait until it holds the messages of all its parents then,
    1) compute the new π value using equation(3.13).
    2) compute the new marginal using equation(3.14).
    3) send a new π_Y message for all its children using the equation(3.6).
  End for
End for

Updating process

When new intervention I is introduced at a given node Y, the updating algorithm will be performed. The added node is a root. The node concerned by the intervention computes its new marginal by combining its previous marginal m_p^Y with the new message π coming from the node DO as follows using equation (3.16). Once the node Y is updated, then it will send new messages to all its neighbors.

- $\pi_{Y \rightarrow Z}$ representing the message sent from a node Y to its children Z using the equation(3.6).

- $\lambda_{Y \rightarrow X}$ representing the message sent from a node Y to its parents using equation(3.7).

Algorithm. Updating
Do pass a message using Equation (3.6) to the node concerned by the intervention (Y)
Define an order of propagation A.
For i from 1 to length(A)
    A(i) computes its new marginal m^Y using equation (3.16) or equation (3.17).
    For i from 1 to length (ne (A(i)))
        Send a new message using equation(3.6) or equation (3.7) to all neighbors not yet been updated.
    End for
End for

Example 3.3 (continued)
An intervention (I_B) is executed upon the augmented graph. A fictive node DO will be added as parent of B. The proposed order is {DO, B, A, D, E, C, F}. The propagation algorithm described above will be performed.

Node DO *Do will pass a message $\pi_{Do \to B}$ to B.*

Node B *When B receives a new π*

- *Compute its new values.*
  So, $m(b_1) = 1 \quad m(b_2) = 0 \quad m(\Theta_B) = 0$
- *Send a new message to all its neighbors (A,D and E): $\pi_{B \to D}$ message to D and $\lambda_{B \to A}$ message to A.*
  $\lambda_{B \to A}(a_1) = 0.85 \qquad \pi_{B \to D}(d_1) = 0.3 \qquad \pi_{B \to E}(e_1) = 0.75$
  $\lambda_{B \to A}(a_2) = 0 \qquad \pi_{B \to D}(d_1) = 0 \qquad \pi_{B \to E}(e_2) = 0$
  $\lambda_{B \to A}(\Theta_A) = 0.15 \qquad \pi_{B \to D}(\Theta_D) = 0.7 \quad \pi_{B \to E}(\Theta_E) = 0.25$

Node A

- *Node A computes its new marginal by combining its previous marginal with this new message π.*
  where $m(a_1) = 0.4 \quad m(a_2) = 0.5 \quad m(\Theta_A) = 0.1$
- *A has no children and no parents not yet been updated.*

Node D

- *Node D computes its new marginal by combining its previous marginal with this new message* $\pi_{B \to D}$.
  where $m(d_1) = 0.27 \quad m(d_2) = 0.26 \quad m(\Theta_D) = 0.47$
- *D has no children and no parents not yet been updated.*

Similarly, we do the same computations for node E, C and F respectively.

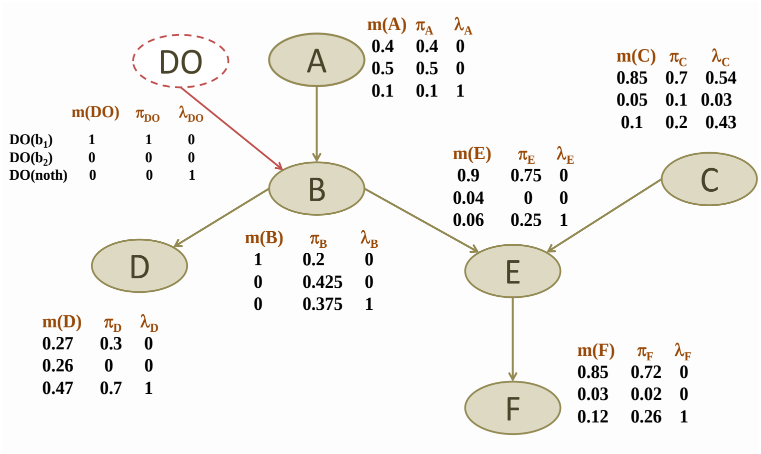

Figure 3.5: The state of the causal belief network upon the intervention $do(b_1)$.

Given an initial causal belief network, acting on a given variable A_i by forcing it to take a specific value a_{ij} amounts to observing the value a_{ij} in its mutilated graph or observing the value $do(a_{ij})$ on the fictive node DO in its associated augmented network.

3.5 Up-down propagation in causal belief networks

An alternative way to make inference in singly connected causal belief network is to implement the up-down belief propagation extended from the standard algorithm of Pearl (1988). This improved version decreases the number of messages since each node is visited at most twice regardless of the number of observed nodes.

The main idea of our algorithm is to propagate information from roots to leaves and then from leaves to roots to ensure that each node receives messages from parents and children. A post-order and a pre-order will be defined to propagate information down and up respectively.

This proposed algorithm is based on four steps:

Step1: Initialization

- Each node updates its values m^X, π_X and λ_X.
- Each node in the post-order receives a message from its parents, updates its π value using equation(3.14) and send a message to its children.

Step2: Updating value of the node concerned by the action
Observations and interventions can be easily handled by augmenting or mutilated the graph, the target variable updates its value using equation(3.15) to deal with observations or equation(3.16) to deal with interventions.

Step3: Propagation-down
Each node sends messages to its children which in turn send messages to their children. Propagation is performed in the direction of leaves following the post-order using equation(3.6).
Each node X in the post-order, combine all messages from its parents and its previous marginal m_p using equation(3.14).

Step4: Propagation-up
Each node sends messages to its parents which in turn send messages to their parents. Propagation is performed in the direction of roots following the pre-order using equation(3.7).
Each node X in the pre-order, combine all messages from its children and its previous marginal m_p using this formula:

$$\lambda_X = m_p \oplus (\oplus_{Y \in Ch(X)} \lambda_{Z \to X}) \tag{3.24}$$

where $\lambda_{Z \to X}$ is computed using equation(3.7).

The phase of initialization is the same for observations and interventions (see section 3.1)

Algorithm. Propagation-down

For i from 1 to length(Post-order)
- send $\pi_{X \to Y}$ message for all children Y of X using equation(3.6).
- for i from 1 to length(Ch(X))
  - Y has to wait until it holds the messages of all its parents then,
  - 1) compute the new marginal using equation(3.14).
  - 2) send a new π_Y message for all its children using equation(3.6).
- End for

End for

Algorithm. Propagation-up

For i from 1 to length(Pre-order)
- send $\lambda_{Y \to X}$ message for all parents X of Y using equation 3.7.
- for i from 1 to length(Pa(X))
  - Y has to wait until it holds the messages of all its children then,
  - 2)compute the new marginal using equation 3..
  - 3)send a new λ_Y message for all its parents using equation 3.7.
- End for

End for

We notice in the following Example 3.4 that the two kinds of propagation give the same results but the second one has a higher cost in the case when we handle one observation.

Example 3.4 *Let us consider the three networks. Network in Figure 3.6 represented the state of the causal network after initialization. The network in Figure 3.7 represented the state of the causal belief network after using the updating phase proposed in our algorithm after an observation $B = b_1$ and the two networks in Figure 4.8 presented the state of the causal belief network using forward and backward propagations.*

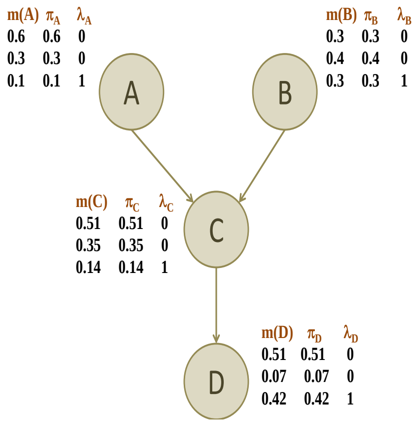

Figure 3.6: The state of the causal belief network after initialization

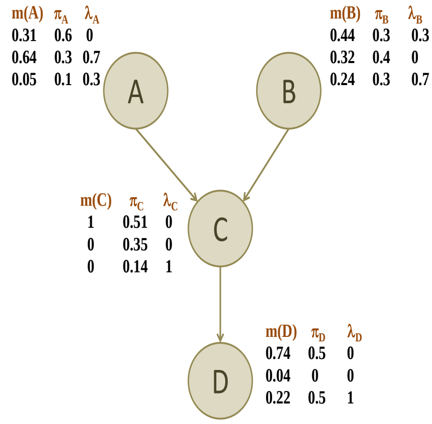

Figure 3.7: The state of the causal belief network after updating

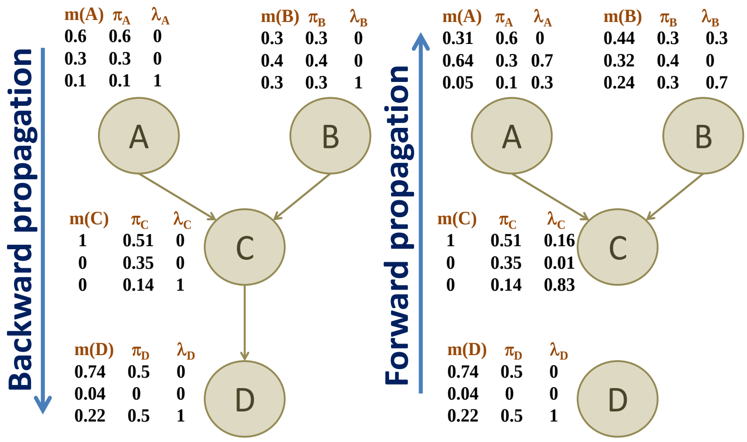

Figure 3.8: The state of the causal belief network after forward and backward propagations

3.6 Conclusion

In this chapter, we have described some algorithms for inference in singly connected networks. Then, we have presented algorithms based on GBT and DRC to propagate information in causal belief networks where distributions are defined per single parents: the centralized belief propagation algorithm the up-down belief propagation algorithm. These algorithms ensure inference in the mutilated graph and in the augmented graph. For that, we have defined conditionals given the "DO" node and given initial causes when beliefs are defined per edge.

In the next chapter, we will deal with the implementation and the experimental simulations for our algorithms.

Chapter 4

Implementation and Simulations

4.1 Introduction

Our goal in this master thesis is to ensure inference in causal belief networks. It consists in computing the effect of a sequence of interventions and observations. This can be done using the graph mutilation method, i.e., by deleting the edges pointing to the node concerned by the action, or using the augmented graph method by adding a new fictive variable. We have proposed in Chapter 3 algorithms allowing the propagation of both observations and interventions in causal belief networks.

In this chapter, we present results relative to the causal belief propagation algorithms. For that, we first investigate on the elapsed time to make inference and then we test if the order in which observations and interventions are introduced is important. After that, in order to explain the usefulness of each proposed algorithm, we will make comparisons between the "centralized" propagation algorithm and the up-down belief algorithm. Finally, we compare between our algorithms and Pearl's algorithms to check if causal belief inference is a generalization of causal Bayesian inference.

This chapter is composed of two parts. First, Section 4.1 deals with tools implementing our algorithms proposed in Chapter 3 for causal inference in singly causal belief networks. Then, Section 4.2 provides results regarding these algorithms.

4.2 Implementation

The implementation of our proposed algorithms allows their evaluation. Experiments ran on a 2.27 GHz Core i3 processor with 4 GO of memory.

Matlab is a high-performance language for numerical computation. It allows one to perform numerical calculations, and visualize the results without the need for complicated and time consuming programming. This can be done in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. It provides a good debugger and profiler which make implementation more efficient. It enables one to write clear and concise code. All these features make Matlab the tool of choice for implementing our algorithm.

We have used some functions of Smets toolbox called TBMLAB implementing under matlab [1].

Our implemented program has as inputs:

- variables
- a priori masses
- conditional masses

This program has as outputs for each variable:

- vector of masses represented the marginal
- π vector represented messages coming from all parents
- λ vector represented messages coming from all children

4.2.1 Generalized Bayesian Theorem (GBT) and Disjunctive Rule of Combination (DRC)

Propagation in causal belief networks obtain their efficiency by making use of the represented independencies between variables in their network. This is done using two rules proposed by Smets and called disjunctive rule of combination (DRC) and generalized Bayesian theorem (GBT) which make possible the direct use of the conditional distribution.

In the following sections, we describe algorithms relative to these rules.

[1] http://iridia.ulb.ac.be/ psmets/

Generalized Bayesian Theorem (GBT)

Let pl_{cond_x} be the set of conditional plausibilities defined given instances of Θ_Y. Suppose m_x is the bba collected at node X. It must be propagated in order to compute GBT_{pl} corresponding to λ_Y vector using equation (3.7). e_x is the cardinality of the node X and e_y is the cardinality of the node Y.

Algorithm 1 Calculate GBT_{pl}

```
DRC_pl ← [ ]
c_x ← 1
for i=1:e_x do
   c_x ← 2 * c_x
end for
c_y ← 1
for i=1:e_y do
   c_y ← 2 * c_y
end for
for i=0:c_y-1 do
   S ← 0
   j ← 1
   cmpt ← i+j
   while j<=length(m_x) do
      S ← S+((m_x(j)) · (pl_cond_x(cmpt)))
      cmpt ← cmpt + c_x*(1);
      j ← j+1
   end while
end for
GBT_pl(length(GBT_pl) + 1) ← S
```

Disjunctive Rule of Combination (DRC)

Let pl_{cond_x} be the set of conditional plausibilities defined given instances of Θ_Y. Suppose m_y is the bba collected at node X. It must be propagated in order to compute DRC_{pl} corresponding to π_X vector using equation (3.6). e_x is the cardinality of the node X and e_y is the cardinality of the node Y.

Algorithm 2 Calculate DRC_{pl}

$DRC_{pl} \leftarrow [\ ]$
$c_x \leftarrow 1$
for i=1:e_x **do**
  $c_x \leftarrow 2 * c_x$
end for
$c_y \leftarrow 1$
for i=1:e_y **do**
  $c_y \leftarrow 2 * c_y$
end for
for i=0:c_x-1 **do**
  S $\leftarrow$ 0
  j $\leftarrow$ 1
  while j<=length(m_y) **do**
    S $\leftarrow$ S+$((m_y(\text{j})) \cdot (pl_{cond_x}((c_x * i) + j)))$
    j $\leftarrow$ j+1
  end while
end for
$DRC_{pl}(length(DRC_{pl}) + 1) \leftarrow S$

4.3 Experimental results

4.3.1 Evaluation criteria

To evaluate our proposed algorithms, we will consider the following parameters:

1. The size complexity: characterized by the number of nodes and number of edges of the causal belief network.

2. The time complexity: characterized by the time taken during the the propagation algorithms.

4.3.2 Up-down algorithm vs centralized algorithm

In Chapter 3, we have proposed two algorithms to deal with sequences, the centralized algorithm and the up-down algorithm. The first one is more suitable in the case of one observation or one intervention. The second allows its efficiency where handling several observations and interventions.

In the following table, we have considered several networks on which we have first acted on a variable, then made an observation followed by an intervention and finally we have handled two observations followed by

two interventions. To compare the performance of the algorithms, we have computed their corresponding running time.
Note that all these networks have binary variables (see Table 4.1).

Table 4.1: Handling observations and interventions (in seconds)

	Centralized belief algorithm			Up-down belief algorithm		
	intervention (observation)	observation + intervention	two observations + two interventions	intervention (observation)	observation + intervention	two observations + two interventions
Network 1	0,031	0,069	0,123	0,069	0,071	0,073
Network 2	0,018	0,037	0,077	0,35	0,035	0,038
Network 3	0,041	0,077	0,142	0,069	0.70	0,072

Using the causal belief propagation algorithm, the number of messages is equal to the number of edges in the augmented graph and in the mutilated graph. So, When we are in front of one observation or one intervention, the execution time during the up-down belief propagation is higher than the centralized propagation because the number of messages circulated in the network is increased. However, when we have one observation and one intervention or two observations (interventions), the number of messages of the two algorithms is almost equal. So, the running time is almost the same. In front of many interventions and observations, the up-down algorithm gives better results regarding the running time.

In the following, we will investigate the running time of our centralized algorithm during the initialization and updating steps.

4.3.3 Running time

Elapsed time during the propagation process

The causal belief propagation algorithm deals with propagation of observations and interventions. It is based on two steps: initialization and updating. Each method wether it is occurred in a time slot. In this section, we will set different cardinalities of the variables and compare time elapsed in initialization and updating phases to propagate an observation and an intervention upon the augmented graph (Gaug) and mutilated graph (Gmut). In fact, this algorithm is shown to be more efficient in case of one intervention or one observation. In the following table, we represent the elapsed time using three networks where all the variables used are binary.

Table 4.2: Elapsed time during observations and interventions (in seconds)

	Observation		Intervention upon the augmented graph		Intervention upon the mutilated graph	
	Initialization	Updating	Initialization	Updating	Initialization	Updating
Network 1	0.022	0,049	0.022	0,058	0.022	0,048
Network 2	0,017	0,022	0,017	0,025	0,017	0,019
Network 3	0,034	0,058	0.034	0,069	0.034	0,057

If we consider a given network, we notice that the initialization process takes the same time for an observation and an intervention since this process is the same. However, the updating process differs and consequently it affects their corresponding running time. The mutilated graph method represents the smaller value since the edges between the node concerned by the action and its parents is deleted. The augmented graph method represents the higher value since adding a fictive node as a parent to the node concerned by the action will increase the number of message-passing.

Elapsed time varying cardinalities

In this section, we are interested in fact of seeing if the running time is much influenced by the cardinality of variables. In the following figures 4.1 and 4.2, we will vary the cardinality of variables for two networks to show the impact of this variation on the value of the running time.

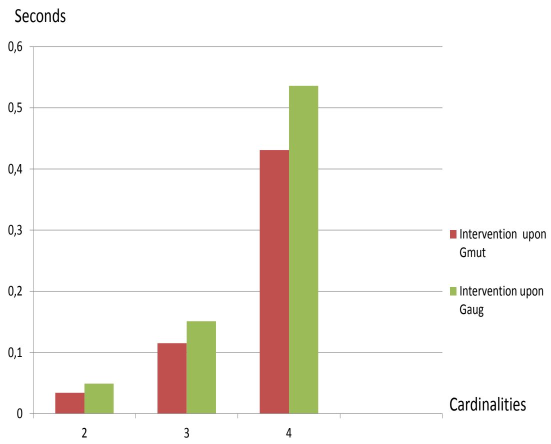

Figure 4.1: Elapsed time by varying cardinalities corresponding to network 1.

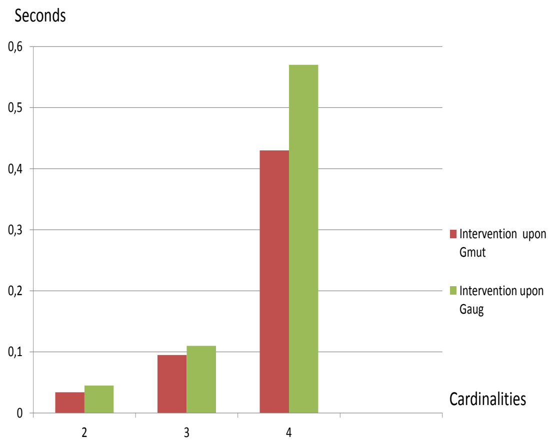

Figure 4.2: Elapsed time by varying cardinalities corresponding to network 2.

As you notice, the cardinality of variables has an important impact on the temporal complexity. In fact, more the size of a variable increases, more tables and a priori conditional distributions become more important.
When we have increased the cardinalities of the variables for the two networks, we have noted a remarkable difference between the running time of each method. This remains valid whatever the used method for handling interventions.

Elapsed time by varying the structure

Varying the structure is based on varying nodes and edges which allows to compute temporal complexity. In this section, we have increased the number of nodes and therefore the number of edges to show if these variations have an impact on the running time of our algorithm (see Table 4.2).

Table 4.3: Elapsed time by varying nodes (in seconds)

Nombre of nodes	Elapsed time during an observation	Elapsed time during an intervention upon Gmut	Elapsed time during an intervention upon Gaug
5	0.04	0.033	0.045
10	0.067	0.061	0.081
15	0.092	0.087	0.102
30	0.225	0.221	0.235

We notice that the number of nodes plays an important role to running time.

4.3.4 Handling both interventions and observations

In this section, we will handle a sequence of observations and interventions given in different order to see the impact from one to the other and the impact of each action on the propagation process. In the following table, we will use two networks represented in Figures 4.3 and 4.4 on which we compute first the effect of a sequence of an observation upon the variable D by seeing d_1 followed by an intervention upon the variable B forcing it to take the value b_1 respectively and then we inverse the order between the intervention and the observation. The effect of these sequences is represented in the following tables 4.4 and 4.5.

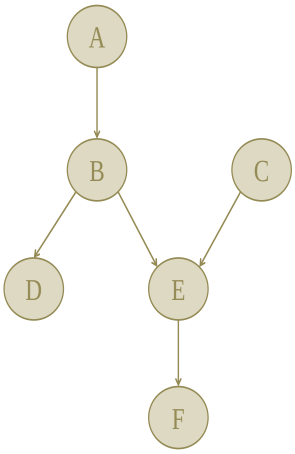

Figure 4.3: Causal belief network 1.

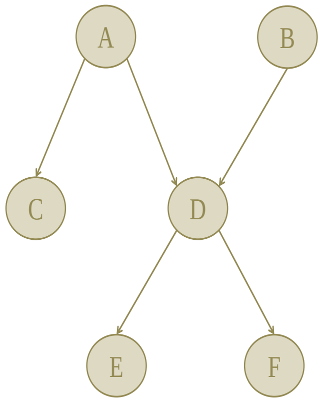

Figure 4.4: Causal belief network 2.

Table 4.4: Handling an observation and an intervention on the network 1.

	Observation followed by Intervention				Intervention followed by Observation			
A	$\emptyset$	a_1	a_2	Θ_A	$\emptyset$	a_1	a_2	Θ_A
	0	0.625	0.281	0.094	0	0.4	0.5	0.1
B	$\emptyset$	b_1	b_2	Θ_B	$\emptyset$	b_1	b_2	Θ_B
	0	1	0	0	0	1	0	0
C	$\emptyset$	c_1	c_2	Θ_C	$\emptyset$	c_1	c_2	Θ_C
	0	0.924	0.035	0.041	0	0.934	0.025	0.041
D	$\emptyset$	d_1	d_2	Θ_D	$\emptyset$	d_1	d_2	Θ_D
	0	1	0	0	0	1	0	0
E	$\emptyset$	e_1	e_2	Θ_E	$\emptyset$	e_1	e_2	Θ_E
	0	0.952	0.023	0.025	0	0.973	0.009	0.018
F	$\emptyset$	f_1	f_2	Θ_F	$\emptyset$	f_1	f_2	Θ_F
	0	0.954	0.012	0.034	0	0.966	0.007	0.027

Table 4.5: Handling an observation and an intervention on the network 2.

	Observation followed by Intervention				Intervention following by Observation			
A	ø	a_1	a_2	Θ_A	ø	a_1	a_2	Θ_A
	0	0.446	0.399	0.155	0	1	0	0
B	ø	b_1	b_2	Θ_B	ø	b_1	b_2	Θ_B
	0	1	0	0	0	0.485	0.365	0.150
C	ø	c_1	c_2	Θ_C	ø	c_1	c_2	Θ_C
	0	0.362	0.444	0.194	0	1	0	0
D	ø	d_1	d_2	Θ_D	ø	d_1	d_2	Θ_D
	0	1	0	0	0	0.407	0.383	0.210
E	ø	e_1	e_2	Θ_E	ø	e_1	e_2	Θ_E
	0	0.304	0.469	0.225	0	0.353	0.420	0.227
F	ø	f_1	f_2	Θ_F	ø	f_1	f_2	Θ_F
	0	0.325	0.461	0.444	0	0.376	0.409	0.215

As shown in Tables 4.4 and 4.5, we should be attentive when we are in front of a sequence of observations and interventions. We notice that these sequences are not equivalent and lead to different results since each action affects differently the network and the inference process. This conclusion is not very surprising. In fact, if we first taste a coffee and observe that it is not hot. Then, after this observation we add ice to the coffee is different when we have added ice and observe that the coffee is not hot.

4.3.5 Inferring a bayesian bba in causal belief network

Since the Bayesian bba (the case where all focal elements are singletons) corresponds to the case of probabilities, we will compare the inference using Bayesian bbas and the inference of the same probabilities using Pearl's algorithm in Bayesian network.

In the algorithm of Pearl, conditional probabilities are defined for all parents. For the special case when conditional beliefs have at most one parent, we find the same results as those found on a Bayesian network.

Example 3.2 *Let us consider the two networks presented in Figure 4.5 and let us suppose that we have an observation* $(b_1 = 1)$ *The network on the right presents the initial bayesian network using conditional probabilities and the other presents the result network after inferring the evidence* b_1.

Therefore, let us consider the two networks presented in the following *figures and we will suppose that we have the same observation. The network on the right represents the initial Bayesian network and the one on the left side represents the network after inferring the evidence b_1.*

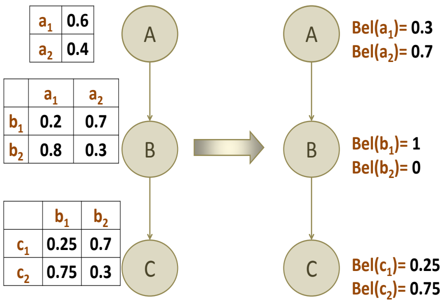

Figure 4.5: Inference in Bayesian networks

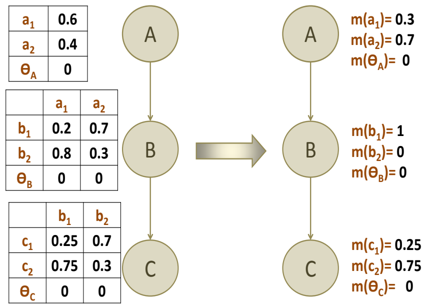

Figure 4.6: Inference in causal belief networks using bayesian bbas

We notice that the belief causal inference using Bayesian bbas and the inference in Bayesian networks using probabilities give the same results.

Concerning the running time of the two algorithms, we have compute the elapsed time for the initial network on which we have made first an observation and then an intervention upon the augmented graph. The results are illustrated in the following Figures.

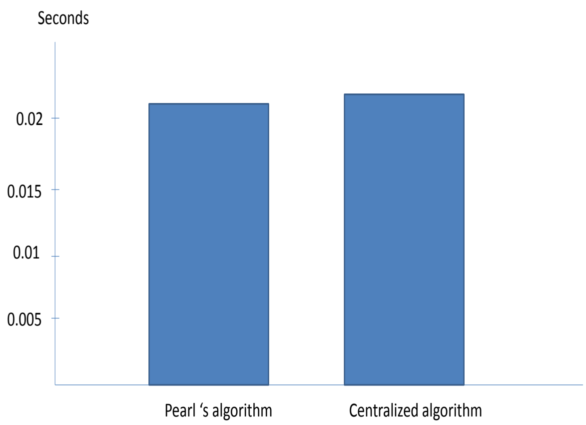

Figure 4.7: Elapsed time on the network 1.

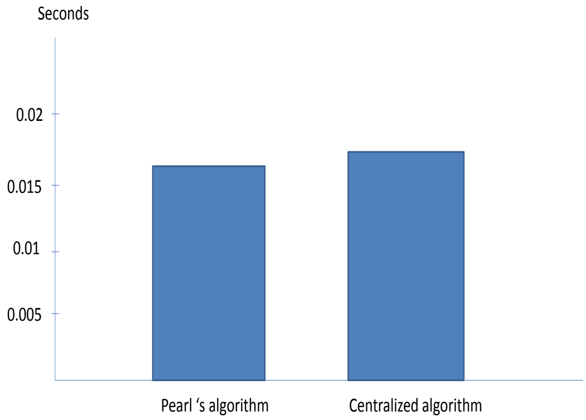

Figure 4.8: Elapsed time on the network 2.

The result of the elapsed time in Bayesian networks is almost similar to our centralized algorithm when bbas are bayesian.
Moreover, this comparison substantiates that the inference in causal belief networks is a generalization of the inference proposed by Pearl since we can model and infer probabilities using our proposed algorithm with conditional bbas.

4.4 Conclusion

In this Chapter, we have presented tools to perform implementation relative to causal belief propagation. Then, we have shown the different results obtained from simulations that have been performed using our different proposed algorithms.

Conclusion

In this master thesis, we are interested in making inference in causal belief networks especially in singly connected ones. In these networks, distributions are defined per single parent.

Since no algorithm deals with causal propagation under the belief function framework, we have proposed algorithms to handle both observations and interventions. The proposed algorithms can be seen as a message-passing scheme for propagating beliefs using local computation based on two rules: the Generalized Bayesian Theorem (GBT) and the Disjunctive Rule of Combination (DRC). These rules make possible the representation of knowledge by conditional distributions.

Since we have considered the case where beliefs are defined per edge, to ensure belief causal inference, we have developed as a first main contribution a new way to define conditional distributions on the augmented graph.

As a second main contribution, we have implemented the "centralized" algorithm which is an adaptation of the algorithm proposed by (Ben Yaghlane & Mellouli, 2008) in context of observational data. Graphically to handle interventions, the structure of the causal network should be altered. Two equivalent changes can be made: The graph mutilation and the graph augmentation method. These algorithms allow to evaluate the effect of interventions and observations. The algorithm is performed in two phases: initialization and updating.

Finally, we have proposed and implemented a new algorithm "up-down" algorithm extended from Pearl's algorithm (Pearl, 1988) and based on these two rules. The main goal of this algorithm is to compute the marginals on all nodes simultaneously. To this end, we must perform the phase of initialization and the two-phases propagation, i.e. propagation-up and propagation-down.

Some interesting future works have to be mentioned. First, we can adapt these algorithms to consider the case where conditional distributions are defined for some parents or for all parents. In our work, we handle standard interventions. Therefore, we can extend our work to take into account non-standards interventions and computes their effects. Note that there is no algorithms dealing with neither interventional nor observational data where beliefs are defined per single parent in multiply connected networks, we can also explore belief causal inference in multi-connected networks where there are loops in the network.

From application point of view, inference in causal belief networks can be used in several applications like those allowing the intrusion detection and or ensuring system reliability. Besides, our approach can also be used in several fields (e.g., marketing, environment).

References

Barnett, J. A. (1981). Computational methods for a mathematical theory of evidence. In *International joint conference on artificial intelligence* (pp. 868–875).

Ben Amor, N., Benferhat, S., & Mellouli, K. (2003). Anytime propagation algorithm for min-based possibilistic graphs. *Soft Computing*, *8*(2), 150-161.

Ben Yaghlane, B., Khalfallah, F., & Mellouli, K. (1998). Propagating conditional belief functions in directed acyclic networks. *In Proceedings of CESA98, IEEE Systems, Man and Cybernetics, volume 2*, 232–238.

Ben Yaghlane, B., & Mellouli, K. (2008). Inference in directed evidential networks based on the transferable belief model. *Int. J. Approx. Reasoning*, *48*(399-418).

Ben Yaghlane, B., Smets, P., & Mellouli, K. (2003). Directed evidential networks with conditional belief functions. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ecsqaru)* (Vol. 2711, p. 291-305). Springer-Verlag.

Benferhat, S. (2010). Interventions and belief change in possibilistic graphical models. *Artificial Intelligence*, *174*(2), 177-189.

Benferhat, S., & Smaoui, S. (2007). Hybrid possibilistic networks. *International Journal of Approximate Reasoning*, *44*(3), 224-243.

Benferhat, S., & Smaoui, S. (2011). Inferring interventions in product-based possibilistic causal networks. *Fuzzy Sets and Systems*, *169*(1), 26-50.

Boukhris, I., Benferhat, S., & Elouedi, Z. (2011). Representing belief function knowledge with directed acyclic graphs. In *International conference on knowledge science, engineering and management* (Vol. 7091, p. 233-245). Springer-Verlag.

Boukhris, I., Benferhat, S., & Elouedi, Z. (2013). Causal belief networks: Handling uncertain interventions. In (pp. 61–72). Springer-Verlag, LNAI 7958.

Boukhris, I., Elouedi, Z., & Benferhat, S. (2011a). Analyzing belief networks with conditional beliefs. In *International conference on intelligent sys-*

tems design and applications (p. 959-964). IEEE Computer Society.

Boukhris, I., Elouedi, Z., & Benferhat, S. (2011b). Modeling interventions using belief causal networks. In *International Florida Artificial Intelligence Research Society Conference (flairs))* (p. 602-607). AAAI Press.

Boukhris, I., Elouedi, Z., & Benferhat, S. (2012). Dealing with interventions with uncertain consequences in belief causal networks. In *International conference on information processing and management of uncertainty in knowledge-based systems* (Vol. 299, p. 585-595). Springer-Verlag.

Boukhris, I., Elouedi, Z., & Benferhat, S. (2013). Dealing with external actions in causal belief networks. In *International journal of approximate reasoning (IJAR), Elsevier, DOI: 10.1016/j.ijar.2013.02.009.*

Cozman, F. (2000). Credal networks. *Artificial Intelligence*, *120*, 199-233.

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks* (Hardcover, Ed.). Cambridge University Press.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multiple valued mapping. *Annals of Mathematical Statistics*, *38*, 325–339.

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, *30*, 205–247.

Denœux, T., & Zouhal, L. M. (2001). Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, *122*, 47–62.

Dubois, D., Grabisch, M., Prade, H., & Smets, P. (2001). Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: the assessment of the value of a candidate. *International Journal of Intelligent Systems*, *16*, 1245–1272.

Dubois, D., & Prade, H. (1986). On the unicity of Dempster rule of combination. *International Journal of Intelligent System*, *1*, 133-142.

Dubois, D., & Prade, H. (2003). Liens causaux et explications. problèmes de modélisation: une discussion préliminaire. *Journées Nationales sur les Modèles de Raisonnement GDR I3 Information-Interaction-Intelligence*, 81-90.

Eberhardt, F., & Scheines, R. (2007). Interventions and causal inference. In *Philosophy of science association* (Vol. 74, p. 981-995).

Elouedi, Z., Mellouli, K., & Smets, P. (2001). Belief decision trees: theoretical foundations. *International Journal of Approximate Reasoning*, *28*(2-3), 91-124.

Goldszmidt, M., & Pearl, J. (1992). Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions. In *International Conference on Principles of Knowledge Representation and Reasoning (kr)* (p. 661-672).

Gordon, J., & Shortliffe, E. H. (1984). The Dempster-Shafer theory of evidence. In *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project* (pp. 272–292).

Halpern, J., & Pearl, J. (2005). Causes and explanations: A structurel model approach. Part i: Causes. *British Journal for the Philosophy of Science*, *56*(45), 843-887.

Hume, D. (2006). *An enquiry concerning human understanding.* Digireads.com.

Jensen, F. (1996). Introduction to bayesian networks. *UCL Press, University college, London.*

Jensen, F., & Nielsen, T. (2007). *Bayesian networks and decision graphs.* Springer.

Kim, J., & Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. *In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI83)*, 190–193.

Kohlas, J., & Monney, P. (1995). A mathematical theory of hints: An approach to Dempster-Shafer theory of evidence. In Springer-Verlag (Ed.), *Lecture notes in economics and mathematical systems.*

Korb, K., Hope, L., Nicholson, A., & Axnick, K. (2004). Varieties of causal intervention. In *Pacific Rim International Conference on Artificial Intelligence (pricai)* (Vol. 3157, p. 322-331). Springer-Verlag.

Lauritzen, S., & Shenoy, P. (1996). Computing marginals using local computation. *Technical Report 267, Kansas University, School of Business.*.

Lauritzen, S., & Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, 50*, 157–224.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* Morgan Kaufmann Pub. San Mateo, Ca, USA.

Pearl, J. (1998). Graphical models for probilistic and causal reasonning. In s. D.M Gabbay P.Smets (Ed.), (Vol. The Handbook of Defeasible Resonning and Uncertainty Management Systems, p. 367-389). Kluwer Academic Press.

Pearl, J. (2000). *Causality: Models, reasonning and inference.* Cambridge University Press.

Pearl, J., & Hopkins, M. (2007). Causality and counterfactuals in the situation calculus. *Journal of Logic and Computation*, *17*, 939-953.

Peot, M. D., & Shachter, D. (1991). Fusion and propogation with multiple observations in belief networks. *Artificial Intelligence 48*, 299-318.

Ristic, B., & Smets, P. (2005). Target identification using belief functions and implication rules. *IEEE Transactions on Aerospace and Electronic Systems*, *41*(3), 1097-1103.

Shachter, R. D. (1988). Probabilistic inference and influence diagrams. *Operations Research, 36*, 589–604.

Shafer, G. (1976). *A mathematical theory of evidence.* Princeton Univ. Press. Princeton, NJ.

Shafer, G. (1986). The combination of evidence. *International journal of intelligent systems*, 155-180.

Shafer, G. (1996). *The art of causal conjecture.* MIT Press.

Shenoy, P. (1993). Valuation networks and conditional independence. In *Uncertainty in artificial intelligence* (p. 191-199).

Simon, C., Weber, P., & Evsukoff, A. (2008). Bayesian networks inference algorithm to implement Dempster-Shafer theory in reliability analysis. *Reliability Engineering and System Safety*, *93*, 950-963.

Smets, P. (1988a). Belief functions. In *Non standard logics for automated reasoning* (pp. 253–286).

Smets, P. (1988b). The transferable belief model for quantified belief representation. In *Handbook of defeasible reasoning and uncertainty management systems* (Vol. 1, pp. 267–301).

Smets, P. (1990). The combination of evidence in the transferable belief model. *IEEE Pattern Analysis and Machine Intelligence*, *12*, 447–458.

Smets, P. (1993a). Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. *Int. J. Approx. Reasoning*, *9*(1), 1-35.

Smets, P. (1993b). Jeffrey's rule of conditioning generalized to belief functions. In *Uncertainty in artificial intelligence* (pp. 500–505).

Smets, P. (1998). The application of the transferable belief model to diagnostic problems. *International journal of intelligent systems*, 127157.

Spirtes, G. C. . S. R., P. (2001). Causation, prediction, and search. *MIT Press*, *1*.

Spohn, W. (1988). Ordinal conditional functions : A dynamic theory of epistemic states. In *Causation in decision, belief changes and statistics* (pp. 105–134).

Teng, C. M. (2012). Applications of causal inference. In *International symposium on artificial intelligence and mathematics.*

Trabelsi, S., Elouedi, Z., & Lingras, P. (2011). Classification systems based on rough sets under the belief function framework. *International Journal of Approximate Reasoning*, *52*(9), 1409-1432.

Verma, T., & Pearl, P. (1990). Equivalence and synthesis of causal models. In *Uncertainty in artificial intelligence* (p. 255-270).

Walley, P. (1991). *Statistical reasoning with imprecise probabilities.* Chapman and Hall, London.

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* Oxford University Press.

Xu, H., & Smets, P. (1994). Evidential reasoning with conditional belief functions. In *Uncertainty in artificial intelligence* (p. 598-606).

Xu, H., & Smets, P. (1996). Reasoning in evidential networks with conditional belief functions. *International Journal of Approximate Reasoning*, *14*, 155–185.