

Université des Sciences et de la Technologie Houari Boumediene

Faculté d'Électronique et d'Informatique

Département d'Informatique

Par:

Hadja Faiza KHELLAF-HANED

**Connaissances imparfaites:
Représentation et Raisonnement**

Année Universitaire 2012-2013

*On mesure l'intelligence d'un individu
la quantité d'incertitude
qu'il est capable de supporter*
EMMANUEL KANT (1724-1804)

Table des matières

Introduction Générale	5
1 Théories de l'incertain et de l'imprécision	7
1.1 Introduction	7
1.2 Quelques notations	7
1.3 La théorie des probabilités	8
1.3.1 Introduction	8
1.3.2 Principes de base	8
1.3.2.1 Mesure de probabilité	8
1.3.2.2 Conditionnement probabiliste	9
1.3.2.3 Théorème de Bayes	10
1.3.3 Indépendance et Indépendance conditionnelle	11
1.3.4 Interprétations des probabilités	12
1.3.5 conclusion	12
1.4 La théorie des fonctions de croyance	13
1.4.1 Fonction de croyance ou de crédibilité	14
1.4.2 Fonction de plausibilité	15
1.4.3 La règle de combinaison de Dempster	16
1.4.4 Le conditionnement	17
1.4.5 Les fonctions de croyances et les probabilités	17
1.4.6 Conclusion sur les fonctions de croyances	18
1.5 La Théorie des possibilités	18
1.5.1 Distribution de possibilités	19
1.5.2 Les mesures de possibilités et de nécessité	19
1.5.3 Conditionnement possibiliste	20
1.5.4 Mesures de probabilités et Mesures de possibilités	21
1.5.5 Mesures de possibilités et les fonctions de croyance	22
1.5.6 Conclusion	22
1.6 Théorie des ensembles flous	23
1.6.1 Fonctions d'appartenance	23
1.6.2 Les caractéristiques d'un ensemble flou	24
1.6.3 Les coupes de niveau	25
1.6.4 Les opérations ensemblistes	25
1.6.4.1 Égalité et inclusion des ensembles flous	25
1.6.4.2 Intersection et Union des ensembles flous	25
1.6.4.3 Complément d'un ensemble flou	26
1.6.5 Mesures de possibilité et de nécessité	27
1.6.6 Raisonnement approché	28
1.6.6.1 L'implication floue	29

1.6.6.2	Le modus ponens généralisé	29
1.6.6.3	Le contrôle flou	30
1.7	Conclusion	30
2	Modèles Logiques de l'incertain	33
2.1	Introduction	33
2.2	La logique possibiliste qualitative	33
2.2.1	Base possibiliste standard	34
2.2.2	Distribution de possibilités associée à une base possibiliste standard	34
2.3	La logique possibiliste quantitative	34
2.4	La logique des pénalités	35
2.4.1	Coût d'une interprétation	36
2.4.2	Coût d'une formule	37
2.5	Les logiques connexes	37
2.5.1	La logique possibiliste et la logique probabiliste	37
2.5.2	La logique possibiliste et la logique modale	38
2.5.3	Raisonnement par défaut utilisant les logiques de l'incertain . . .	38
2.5.4	La logique des pénalités et la logique possibiliste	39
2.5.5	La logique des pénalités et les fonctions de croyance	40
2.5.6	La logique des pénalités et les fonctions conditionnelles ordinales	40
2.6	Conclusion	41
3	Modèles Graphiques de l'incertain : Représentation et Raisonnement	43
3.1	Introduction	43
3.2	Réseaux causaux Bayésiens	44
3.2.1	Les concepts de base	44
3.2.1.1	Quelques définitions	45
3.2.1.2	Les réseaux causaux et le critère de D-Séparation	46
3.2.1.3	L'indépendance conditionnelle dans les DAGs	50
3.2.1.4	Modélisation d'un réseau Bayésien	50
3.2.2	Algorithmes de propagation dans les réseaux causaux bayésiens .	52
3.2.2.1	Introduction	52
3.2.2.2	La propagation dans les Polyarbres	53
3.2.2.3	La propagation dans les graphes à connexions multiples	56
3.3	Réseaux causaux possibilistes	67
3.3.1	Définition d'un réseau causal possibiliste	68
3.3.1.1	Les réseaux causaux possibilistes basés sur le produit :	68
3.3.1.2	Les réseaux causaux possibilistes basés sur le minimum :	69
3.3.2	Propagation dans les réseaux causaux possibilistes basés sur le	
minimum		70
3.3.2.1	Propagation basée sur le minimum dans les polyarbres .	70
3.3.2.2	Propagation basée sur le minimum dans les graphes à	
connexions multiples		72
3.4	Réseaux causaux basés sur le produit	75
3.5	Propagation basée sur le produit dans les polyarbres	77
3.5.1	Algorithme de propagation :	78
3.6	Propagation basée sur le produit dans les réseaux causaux possibilistes à	
connexions multiples		79
3.6.1	Algorithme de la propagation	80
3.7	Conclusion	83

<i>Table des matières</i>	3
4 L'inférence en logique possibiliste	85
4.1 Introduction	85
4.2 L'inférence en logique possibiliste standard (SPL)	85
4.3 L'inférence en logique possibiliste quantitative	88
4.3.1 Le problème WMAXSAT	88
4.3.2 Principe de l'algorithme	88
4.3.3 Complexité de l'algorithme d'inférence en logique des pénalités	91
Bibliographie	97

Introduction Générale

L'intelligence artificielle a pour vocation de développer des outils pour représenter les connaissances et de fournir des mécanismes d'inférence pour élaborer des conclusions à partir des connaissances disponibles [1].

C'est la capacité de transmettre des représentations qui a permis l'extraordinaire accélération des possibilités humaines [2]. Quant à l'inférence, c'est l'élément de base de toute description de la cognition [2].

Néanmoins, les connaissances du monde dont nous disposons sont imparfaites. Ceci est dû à plusieurs facteurs liés soit à la manière dont nous récoltons ces connaissances soit à la nature même de ces connaissances.

L'imperfection est perçue soit comme une incertitude soit comme une imprécision.

La notion d'incertitude est une forme subjective de l'ignorance [3]. Elle apparaît quand un agent construit une opinion subjective sur une proposition qui n'est pas définitivement établie.

L'imprécision est une forme objective de l'ignorance. Elle couvre les cas où un agent attribue une valeur à une variable mais sans la précision adéquate. Pour illustrer la différence entre imprécision et incertitude, considérons les deux situations suivantes [4] :

1. Mohamed a au plus deux enfants et j'en suis sûr.
2. Mohamed a trois enfants mais je n'en suis pas sûr.

Dans le premier cas, le nombre d'enfants est imprécis mais certain. Dans le second cas, le nombre d'enfants est précis mais incertain. Ces deux aspects (imprécision et incertitude) peuvent coexister mais ils sont distincts.

Durant plusieurs décennies, les différents aspects de l'imperfection de l'incertitude ont été modélisés par la théorie des probabilités. La notion de probabilité n'est cependant pas adaptée à la représentation des incertitudes de nature psychologiques [5], telles que celles liées à la fiabilité d'un informateur. Ainsi, durant ces trente années, plusieurs modèles ont émergé, reflétant les nouveaux concepts prenant en compte les différents aspects de l'imperfection, confirmant ainsi que la théorie des probabilités, aussi répandue qu'elle l'est, n'est pas le seul modèle normatif pouvant capturer l'ensemble de ces aspects [4].

La théorie de l'évidence [6] permet également de manipuler les degrés de confiance qu'un observateur attribue à la validité des faits, en n'imposant plus la condition d'additivité fondamentale de la théorie des probabilités.

Afin de modéliser la notion d'imprécision, Zadeh a proposé de modéliser le mécanisme de la pensée humaine par un raisonnement approximatif basé sur des variables linguistiques. Il a introduit alors les ensembles flous [7]. Et afin de tenir compte du caractère incertain et non probabilisé des connaissances, Zadeh [8] a introduit une théorie non probabiliste représentée par la théorie des possibilités, qui fût par la suite développée par Dubois et Prade [9].

Au fil de l'évolution des recherches, d'autres types de représentation ont vu le jour comme les représentations graphiques. Les réseaux bayésiens probabilistes [10, 11, 12] figurent parmi les modèles graphiques les plus répandus. Ils mettent en exergue les relations d'indépendances qui existent entre les variables.

La représentation et le raisonnement sur des connaissances incertaines en théorie des possibilités, peuvent être abordés dans des cadres symbolique ou numérique.

Dans le cadre symbolique, une relation d'ordre est définie entre les formules constituant les connaissances. Nous parlons alors de la théorie des possibilités qualitative.

Par contre dans le contexte numérique, un degré numérique, évaluant l'incertitude, est associé à chaque parcelle de connaissances. Ce cas correspond à la théorie des possibilités quantitative.

Pour chaque contexte, il existe deux façons de codifier les connaissances :

- soit en utilisant le mode logique où les connaissances incertaines sont représentées sous forme de bases de connaissances formées par un ensemble de formules propositionnelles associées à des pondérations mesurant l'incertitude,
- soit en utilisant le mode graphique où les connaissances sont représentées par un certain nombre de variables représentées par des nœuds et un ensemble d'arcs représentant les liens causaux qui existent entre les différentes variables. Au niveau de chaque nœud est attribuée une valuation exprimant l'influence des incertitudes entre les variables dépendantes.

Le but de ce cours est de donner un aperçu sur les principales théories qui traitent les connaissances imparfaites et plus précisément les connaissances incertaines.

Chapitre 1

Théories de l'incertain et de l'imprécision

1.1 Introduction

Plusieurs modèles numériques ont été établis pour traiter ces différents types d'ignorance. La théorie des probabilités fût la première à s'être intéressée à la modélisation de l'incertitude. En effet, les incertitudes ont été identifiées depuis fort longtemps par l'intermédiaire de la théorie des probabilités [5]. Elle permet, par exemple, de quantifier la croyance pour qu'une proposition soit vraie. Cependant, elle est assez limitée pour exprimer certaines formes d'incertitudes [3].

Par la suite, les travaux de Dempster et Shafer sur la théorie des fonctions de croyance sont apparus afin de modéliser la notion de crédibilité issue des connaissances du type "Mon degré de croyance que le cancer X est dû à un virus". Vu que les connaissances incertaines peuvent ne pas être de nature probabiliste (numérique), la théorie des possibilités fût introduite afin de représenter des possibilités épistémiques du type "la possibilité que la hauteur de Paul soit autour de 175 cm".

Le concept de l'imprécision fût modélisé par les ensembles flous [13].

Nous allons nous focaliser, dans ce chapitre, sur ces différentes approches numériques proposées pour représenter les connaissances incertaines et imprécises.

1.2 Quelques notations

Nous allons dans cette section présenter les conventions de notations utilisées dans cette thèse.

Soit \mathcal{L} le langage propositionnel. Ω représente l'ensemble des interprétations propositionnelles. ω représente une interprétation de Ω . ϕ, ψ, \dots représentent des formules propositionnelles. A, B, C, \dots représentent des variables. \top représente une tautologie, \perp représente une antilogie et \neg représente la négation, \wedge représente la conjonction et \vee représente la disjonction.

Soit $\omega \in \Omega$ une interprétation et soit ϕ une formule propositionnelle. $\omega \models \phi$ signifie que l'interprétation ω est un modèle de ϕ dans le sens de la logique propositionnelle. $\omega \not\models \phi$ signifie que l'interprétation ω n'est pas un modèle de ϕ .

1.3 La théorie des probabilités

1.3.1 Introduction

De toutes les théories qui représentent l'incertitude, la théorie des probabilités [10] est la plus ancienne et la plus répandue. Elle constitue une base solide pour le raisonnement logique avec incertitude car elle permet de modéliser quantitativement l'incertitude associée à un événement.

Considérons par exemple l'événement N "*il neigera sur les hauts plateaux d'ici la fin de semaine*". L'occurrence de N est difficile à prédire. Cette incertitude est quantifiée par un nombre $P(N)$, appelé la probabilité associée à N. Cette probabilité est non négative et ne doit pas excéder 1.

1.3.2 Principes de base

Les bases de la formalisation axiomatique de la théorie des probabilités sont les suivantes :

1.3.2.1 Mesure de probabilité

La notion de base de la théorie des probabilités est la mesure de probabilités définie dans un espace de probabilité.

Définition 1.1 (Espace de probabilités) (*Kolmogoroff, 1933*) *Un espace de probabilités formel est défini par trois entités (S, F, P) où :*

1. S est l'espace de réalisation,
2. F est l'ensemble des événements en S ,
3. P est une mesure de probabilité définie sur les éléments de F .

Sur l'espace de probabilités S , sont définies des opérations algébriques entre événements :

- \bar{A} représente l'événement complémentaire de A ,
- $A \cup B$ représente l'union de deux événements A et B ,
- $A \cap B$ représente l'intersection de deux événements A et B ,
- $A \setminus B$ représente la différence entre deux événements A et B .

Considérons deux événements A et B . Deux cas particuliers sont à noter :

- A et B sont dits exhaustifs si $A \cup B = S$.
- A et B sont dits exclusifs si $A \cap B = \emptyset$ où \emptyset est l'événement impossible.

La mesure de probabilités P satisfait les trois axiomes de base suivants :

1. $P(\emptyset) = 0$;
2. $P(S) = 1$;
3. Si A_1, A_2, \dots, A_n sont des événements exclusifs ($A_i \cap A_j = \emptyset \forall i \neq j$), alors :

$$P(\cup_{h=1}^n A_h) = \sum_{h=1}^n P(A_h).$$

Le troisième axiome est l'axiome d'additivité propre au modèle probabiliste dans le cas où les A_i sont mutuellement exclusifs. Il atteste que la croyance d'un événement est la somme des croyances assignées à ces composants exclusifs.

A partir de ces axiomes sur la fonction de probabilité P , les propriétés élémentaires suivantes sont dérivées :

- Propriété 1.1**
1. $\forall A \in S, 0 \leq P(A) \leq 1$. Les deux valeurs extrêmes sont interprétées comme suit : $P(A)=0$ signifie que l'événement A est impossible et $P(A)=1$ signifie que A est certain.
 2. $\forall A \in S, P(A)+P(\bar{A})=1$. C'est une équation qui permet de déduire la probabilité d'un événement dès que la probabilité de l'événement contraire est connue. Ceci implique que quand nous augmentons la croyance d'un événement, nous diminuons la croyance de l'événement contraire.
 3. Soient A et B deux événements, la réalisation de A et B quand A et B sont impossibles est un événement impossible :

$$A \cap B = \emptyset \Rightarrow P(A, B) = 0$$

4. Soient A et B deux événements.
Si $A \subseteq B$ Alors $P(A) \leq P(B)$.
Si $A \subset B$ alors $P(B \setminus A) = P(B) - P(A)$.
5. Soient A et B deux événements.

$$P(A \cup B) + P(A, B) = P(A) + P(B) \Leftrightarrow P(A \cup B) = P(A) + P(B) - P(A, B)$$

Lorsque A et B sont exclusifs auquel cas, $P(A \cap B) = 0$, alors, $P(A \cup B) = P(A) + P(B)$.
Plus généralement, si tous les A_i sont incompatibles deux à deux alors :

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$$

1.3.2.2 Conditionnement probabiliste

La notion du conditionnement est inhérente à la théorie des probabilités. Elle permet de modéliser les liens de dépendance entre deux événements. Plus précisément, elle évalue l'impact de l'arrivée d'une information certaine sur les connaissances du monde.

Exemple 1.1 Soit l'expérience de jeter un dé. Soit X un jet, F l'événement $\{X=6\}$ et E l'événement $\{X>4\}$. Ainsi, $P(F)=1/6$. Supposons qu'un dé a été jeté et que l'événement E s'est produit générant ainsi deux possibilités : (5 ou 6). Intuitivement, en l'absence d'autres informations : $P(F | E)=1/2$.

$P(F | E)$ est appelée probabilité conditionnelle de F sachant E .

D'une manière formelle, le conditionnement est donc matérialisé par la définition suivante :

Définition 1.2 (Probabilité Conditionnelle) Soient A et B deux événements avec $P(B)>0$. La probabilité conditionnelle de A sachant B est définie par :

$$P(A | B) = \frac{P(A, B)}{P(B)} \quad (1.1)$$

$P(B)$ est dite probabilité marginale.

Exemple 1.2 Si on reprend l'exemple précédent du dé avec F l'événement $\{X=6\}$ et E l'événement $\{X>4\}$. L'événement $E \cap F$ est l'événement F d'où :

$P(F | E) = \frac{P(F \cap E)}{P(E)} = \frac{P(F)}{P(E)} = \frac{1/6}{1/3} = 1/2$. Ce qui correspond au résultat obtenu précédemment.

Propriété 1.2 1. Si $A \supset B \Rightarrow P(A | B) = 1$ (car $A \cap B = B$).

2. Si $A \subset B \Rightarrow P(A | B) \geq P(A)$ (car $A \cap B = A$).

3. Les probabilités conditionnelles satisfont les axiomes de la mesure de probabilité.

A partir de la définition de la probabilité conditionnelle, est déduite la probabilité de l'intersection de deux événements, appelée probabilité jointe, en termes de probabilité conditionnelle et de probabilité marginale comme suit :

$$P(A \cap B) = P(A | B) * P(B) = P(B | A) * P(A) \quad (1.2)$$

Cette règle peut être appliquée à plusieurs d'événements produisant ainsi la règle de multiplication ou la règle de chaînage.

Définition 1.3 (Règle de chaînage) Soient A_1, A_2, \dots, A_n n événements, la probabilité jointe associée aux événements (A_1, A_2, \dots, A_n) est définie comme le produit des n probabilités conditionnelles comme suit :

$$P(A_1, A_2, \dots, A_n) = P(A_1 | A_2, \dots, A_{n-1}, A_n) * \dots * P(A_{n-1} | A_n) * P(A_n) \quad (1.3)$$

A partir de l'équation des probabilités jointes et de l'axiome 3 de la mesure des probabilités, la notion de probabilité totale est définie par : Soit $\{A_i\}$ une partition de S , c'est à dire :

- $A_i \cap A_j = \emptyset, i \neq j$
- $S = \bigcup_{i=1, \dots, N} A_i$

Alors, pour tout événement B de S ,

$$P(B) = \sum_{i=1}^N P(B, A_i) = \sum_{i=1}^N P(B | A_i) * P(A_i)$$

Cette définition exprime la probabilité marginale d'un événement B comme une moyenne pondérée des probabilités conditionnelles.

La règle de chaînage et le théorème de la probabilité totale peuvent être étendus aux probabilités conditionnelles :

$$P(A_1, A_2, \dots, A_n | C) = P(A_n | A_{n-1}, A_{n-2}, \dots, A_1, C) * \dots * P(A_2 | A_1, C) * P(A_1 | C) \quad (1.4)$$

$$P(B | C) = \sum_{i=1}^N P(B | A_i, C) * P(A_i, C) \quad (1.5)$$

1.3.2.3 Théorème de Bayes

Le théorème de Bayes est une formulation mathématique du conditionnement modélisant la modification des croyances après l'arrivée d'une nouvelle information liant les probabilités conditionnelles de deux événements A et B . Il stipule :

$$P(A | B) = \frac{P(B | A) * P(A)}{P(B)} \quad (1.6)$$

avec $P(A) \neq 0$ et $P(B) \neq 0$.

Le théorème de Bayes peut être vu comme une règle qui permet de mettre à jour une probabilité initiale $P(A)$ appelée aussi probabilité a priori donnant ainsi lieu à une probabilité révisée $P(A|B)$, appelée probabilité a posteriori.

La règle de Bayes joue un rôle important pour toutes sortes d'applications comme pour le diagnostic médical, car elle permet de calculer la probabilité d'un diagnostic à partir de la probabilité de la cause :

$$P(\text{cause} | \text{effet}) = \frac{P(\text{effet} | \text{cause})P(\text{cause})}{P(\text{effet})} \quad (1.7)$$

Cependant, la difficulté d'un problème augmente infiniment lorsque les variables ne sont pas indépendantes. Ainsi, l'utilisation du théorème de Bayes, nécessite la prise en compte des indépendances conditionnelles.

1.3.3 Indépendance et Indépendance conditionnelle

Deux événements A et B sont indépendants si leurs probabilités ne changent pas quand nous les conditionnons l'un à l'autre : $P(A|B)=P(A)$ et $P(B|A)=P(B)$ avec $P(A)>0$ et $P(B)>0$. Cette indépendance peut être également formalisée par :

Définition 1.4 (Indépendance) Deux événements A et B sont dits indépendants, avec $P(A) > 0$ et $P(B) > 0$, si $P(A,B)=P(A)*P(B)$

En d'autres termes, la probabilité jointe de ces deux événements correspond au produit des probabilités marginales. Si A et B sont indépendants, alors $(A \text{ et } \bar{B})$, $(\bar{A} \text{ et } B)$ et $(\bar{A} \text{ et } \bar{B})$ le sont également.

La notion d'indépendance peut être étendue à un ensemble fini d'événements A_1, A_2, \dots, A_n .

Définition 1.5 Un ensemble d'événements A_1, A_2, \dots, A_n est dit indépendant si pour chaque sous ensemble A_i, A_j, \dots, A_m de ces événements, nous avons :

$$P(A_i, A_j, \dots, A_m) = P(A_i) * P(A_j) * \dots * P(A_m)$$

Remarque 1.1 Il est important de remarquer que la notion d'indépendance ne coïncide pas avec celle d'événements mutuellement exclusifs. En effet, si A et B sont mutuellement exclusifs, alors leur intersection est vide et donc $P(A,B)=0$, ce qui entraîne $P(A|B)=0$. Ainsi, si B se produit, l'occurrence de A est impossible.

Définition 1.6 (Indépendance conditionnelle) Deux événements A et B, avec $P(A)>0$ et $P(B) > 0$ sont dits indépendants dans le contexte C si :

$$P(A, B | C) = P(A | C) * P(B | C)$$

Il est important de remarquer que l'indépendance conditionnelle n'implique pas l'indépendance. En effet, deux événements A et B peuvent être indépendants dans le contexte C mais si nous les considérons à part, ils ne seront pas indépendants.

Remarque 1.2 Il est important à ce niveau de souligner l'impact de la notion d'indépendance sur la complexité du raisonnement probabiliste. En effet, le calcul des probabilités requiert :

- Soit la connaissance d'un nombre important de probabilités conditionnelles du type $P(A|B)$,

- Soit l'exploitation des indépendances pour un certain nombre de couples d'événements (A, B) .

Ainsi, l'utilisation de l'indépendance probabiliste entre certains événements permet de réduire notablement la complexité des calculs pour le raisonnement probabiliste dans l'incertain [14]. En effet, la difficulté d'un problème augmente considérablement lorsque les variables ne sont pas indépendantes.

1.3.4 Interprétations des probabilités

Le modèle probabiliste peut être interprété de plusieurs manières. Essentiellement, nous parlons des probabilités fréquentistes ou des probabilités subjectives.

Selon l'interprétation fréquentiste, les probabilités sont obtenues à partir de la notion de "fréquence relative d'apparition lors de la répétition d'un grand nombre de fois d'une expérience aléatoire".

Exemple 1.3 Soit la situation où il s'agit de tirer une boule au hasard dans un ensemble de n boules formé de r boules rouges et de j boules jaunes. La probabilité que la proposition $A =$ "La boule tirée est rouge" correspond à la fréquence relative avec laquelle nous nous attendons à tirer une boule rouge. $P(A) = r/n$.

L'interprétation fréquentiste ne permet pas de raisonner sur l'incertain si les données statistiques disponibles ne sont pas suffisantes, ou si le cadre d'une expérience aléatoire répétée dans des conditions identiques n'est pas respecté [14].

L'approche subjective peut être définie par le slogan "La probabilité est un degré de croyance". En effet, il y a des situations pour lesquelles il n'y a pas d'interprétations fréquentistes objectives comme par exemple la probabilité d'avoir une bonne note à un examen.

1.3.5 conclusion

Le modèle probabiliste reste le modèle le plus utilisé. Néanmoins, il présente des limites :

- La théorie des probabilités ne permet pas de distinguer entre l'ignorance et l'incertitude, qui cependant, dans le sens commun, ne recèlent pas la même quantité de savoir. En effet, soient A et B deux événements. Affirmer que $P(A) = P(B) = 0.5$ peut désigner l'équiprobabilité entre les deux événements ou encore la totale ignorance sur les événements A et B . En effet, l'axiome d'additivité de la théorie des probabilités entraîne que si nous ne savons rien sur un événement A alors

$$P(A) = P(\neg A) = 1/2.$$

Le problème majeur qui en découle repose sur le fait que d'augmenter les croyances sur un événement entraîne la diminution sur la croyance de l'événement contraire, ce qui n'est pas toujours souhaitable.

- De plus, il est difficile en théorie des probabilités de représenter l'ignorance totale lorsqu'on se trouve en présence de plus de deux alternatives. En effet, considérons le paradoxe suivant [15].

Exemple 1.4 Soient les propositions $P1 =$ "La vie existe sur Mars", $P2 = \neg P1 \wedge P3 =$ "La vie sur Mars est seulement végétale", $P4 =$ "La vie sur Mars n'est pas seulement végétale"; Il est clair que $P3 \wedge P4 = \emptyset$ et que $P1 = \neg P3 \vee P4$. Quelle que soit la répartition des probabilités (par exemple $P2 = 1/2$, $P3 = P4 = 1/4$ ou

$P2=P3=P4=1/3$) utilisée pour rendre compte de notre ignorance a priori, certaines propositions apparaissent plus probables que d'autres, ce qui est paradoxal vu qu'on est dans le cas d'une ignorance totale.

Ces problèmes découlent de l'axiome d'additivité.

- Un autre inconvénient du calcul des probabilités réside dans la nécessité d'énumérer un ensemble exhaustif d'alternatives mutuellement exclusives [16]. En pratique, l'ensemble des faits fournis par un expert ne répondent pas forcément à ces exigences.
- L'utilisation du calcul des probabilités conduit souvent, en pratique, à poser des axiomes d'indépendance entre les faits manipulés [16].
- Du point de vue du calcul, les règles de calcul de la théorie des probabilités favorisent l'augmentation des erreurs (il y a des sommes et des produits). Si bien qu'une longue chaîne de propagation des coefficients d'incertitude peut réduire à néant la signification de ces derniers [16].

Pour tout cela, des travaux de recherche ont conduit à d'autres formalismes pour représenter l'incertain.

1.4 La théorie des fonctions de croyance

Pour pallier aux problèmes posés par le modèle probabiliste, notamment ceux liés à l'axiome d'additivité, Dempster et Shafer [6] ont défini la notion de fonctions de croyances qui ont pour objectif de modéliser et de quantifier la notion de degrés de croyance pour représenter l'incertitude. Dans la théorie des croyances [17, 18, 19], les distributions de probabilités sont construites non pas sur Ω mais sur l'ensemble des parties de Ω . Ω est dit cadre de discernement représentant l'ensemble exhaustif et exclusif des états connus du monde. 2^Ω est l'ensemble des parties de Ω . La distribution représente la masse de probabilités m définie par :

$$m : 2^\Omega \rightarrow [0,1]$$

Les propriétés de la fonction de masse sont :

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad (1.8)$$

$$m(\emptyset) = 0. \quad (1.9)$$

L'égalité (1.8) représente la condition de normalisation.

Tout sous-ensemble de Ω vérifiant $m(A) \neq 0$ est dit élément focal de m .

Le couple (Ω, m) est appelé corps de la croyance.

Si $m(A)=1$ et $\forall B \neq A \in 2^\Omega, m(B)=0$, alors A est certain dans le sens où un de ces éléments a la valeur cherchée. Dans ce cas nous ne savons pas quel est l'élément concerné sauf si A est un singleton.

Si $A=\Omega$, alors l'information $m(A)=1$ correspond à une situation d'ignorance totale.

Exemple 1.5 [20] *Imaginons que nous nous intéressons à l'évolution sur 24 heures du cours d'une action en bourse. Ainsi, au bout de 24 heures, soit le cours aura changé (événement a_1), soit il restera stable (événement a_2). Il faut attribuer une masse à chacun des sous-ensembles possibles : $m(\emptyset), m(\{a_1\}), m(\{a_2\}), m(\{a_1, a_2\})$. Supposons que selon une analyse financière, le cours restera stable, mais sans grande certitude. Nous en déduisons que $m(\{a_2\})=0.3$. D'un autre côté, le fait que le marché boursier est*

très volatil, permet de dire que le cours aura tendance à fluctuer. Nous en déduisons que $m(\{a_1\})=0.2$. La masse restante sera attribuée à $\{a_1, a_2\}$ qui se traduit par "Nous ne pouvons pas nous prononcer". Il s'agit de la masse restante qui n'a pas encore été attribuée. En résumé, notre crédibilité sera partagée entre une partie de connaissance et une partie d'ignorance. Nous aurons :

- $m(\emptyset)=0$ par définition,
- $m(\{a_1\})=0.2$,
- $m(\{a_2\})=0.3$,
- $m(\{a_1, a_2\})=0.5$.

L'élément focal correspond à un événement en lequel croit un observateur au minimum un peu. L'ensemble des éléments focaux constitue le noyau : $N_\Omega = \{A \in 2^\Omega : m(A) \geq 0\}$.

1.4.1 Fonction de croyance ou de crédibilité

A la fonction de masse m est associée la notion de fonction de croyance Bel définie comme suit : la fonction de croyance évalue l'intensité de l'implication d'une formule à partir des croyances disponibles. Ainsi, si A est un ensemble dans un cadre de discernement Ω , $Bel(A)$ représente la somme de toutes les masses attribuées aux éléments contenus dans A , plus la masse attribuée à A lui même. D'une manière plus formelle, la fonction de croyance est définie par : $Bel : 2^\Omega \rightarrow [0,1]$ telle que :

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (1.10)$$

La fonction de crédibilité vérifient les propriétés suivantes :

- $Bel(\emptyset)=0$
- $Bel(\Omega)=1$
- Règle de sous-additivité :

$$Bel(A) + Bel(\bar{A}) \leq 1$$

Cette propriété permet d'augmenter la croyance d'un événement sans pour autant diminuer celle de l'événement contraire. Elle offre une souplesse qui fait défaut à la théorie des probabilités à cause de l'axiome d'additivité qui est contraignant.

-

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B).$$

Concernant cette propriété, la distinction est importante. En effet, pour la théorie des probabilités, le degré de probabilité accordé à l'union de deux événements mutuellement exclusifs est égal à la somme des probabilités allouées à chaque événement.

- Règle de monotonie :

$$Si A \subseteq B \Rightarrow Bel(A) \leq Bel(B).$$

Exemple 1.6 Si nous reprenons l'exemple 1.5, nous avons les masses suivantes : $m(\emptyset)=0$, $m(\{a_1\})=0.2$, $m(\{a_2\})=0.3$, $m(\{a_1, a_2\})=0.5$. Les degrés de crédibilité qui en résultent sont :

- $Bel(\emptyset)=0$,
- $Bel(\{a_1\})=0.2$,
- $Bel(\{a_2\})=0.3$,

- $Bel(\{a_1, a_2\})=1$ car la somme des masses de tous les sous-ensembles du cadre de discernement est toujours égale à 1 ($\Omega=\{a_1, a_2\}$).

La fonction de crédibilité est également considérée comme un élément focal. L'ensemble de tous les éléments focaux d'une fonction de crédibilité s'appelle le cœur de la fonction de crédibilité.

1.4.2 Fonction de plausibilité

La fonction de plausibilité évalue dans quelle mesure un événement est consistant avec les croyances disponibles. Elle est définie par : $Pl : 2^\Omega \rightarrow [0,1]$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (1.11)$$

La plausibilité d'un ensemble d'éléments A dans un cadre de discernement Ω est la somme des croyances qui ne contredisent pas A.

La fonction de plausibilité peut également être définie à travers la relation suivante :

$$Pl(A) = 1 - Bel(\bar{A})$$

En effet, le degré de plausibilité de A représente tout le degré de crédibilité qui n'est pas attribué à \bar{A} . En d'autres termes, A n'est quelque peu crédible que dans la mesure où l'événement opposé n'est pas très plausible. La fonction de plausibilité satisfait les relations suivantes :

- $Pl(\emptyset)=0$
- $Pl(\Omega)=1$
- $Pl(A)=1-Bel(\bar{A})$
- $Pl(A)+Pl(\neg A) \geq 1$
- Règle de monotonie : Si $A \subseteq B$ Alors $Pl(A) \leq Pl(B)$
- $Pl(A \cap B) \leq Pl(A)+Pl(B)-Pl(A \cup B)$

Exemple 1.7 Si nous reprenons l'exemple 1.5, nous avons les masses suivantes : $m(\emptyset)=0$, $m(\{a_1\})=0.2$, $m(\{a_2\})=0.3$, $m(\{a_1, a_2\})=0.5$. Les degrés de plausibilité qui en résultent sont :

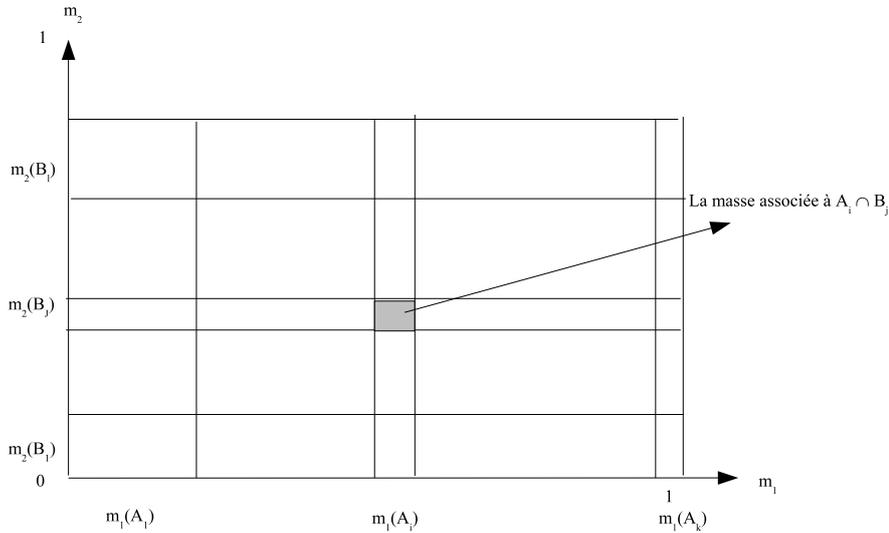
- $Pl(\emptyset)=0$,
- $Pl(a_1)=m(\{a_1\})+m(\{a_1, a_2\})=0.7$,
- $Pl(a_2)=m(\{a_2\})+m(\{a_1, a_2\})=0.8$,
- $Pl(\{a_1, a_2\})=Pl(\Omega)=1-Bel(\emptyset)=1$ en utilisant la relation entre la fonction de crédibilité et la fonction de plausibilité.

La fonction de crédibilité et la fonction de plausibilité sont reliées par plusieurs points :

- Si nous comparons la fonction de crédibilité avec la fonction de plausibilité, il apparaît que la fonction de plausibilité domine la fonction de crédibilité. Ceci est illustré par l'équation suivante :

$$\forall A \subseteq B, Bel(A) \leq Pl(A)$$

- $\forall A \subseteq \Omega, Bel(A)+Bel(\bar{A}) \leq 1$ et $Pl(A)+Pl(\bar{A}) \geq 1$.

FIGURE 1.1 – Somme orthogonale entre m_1 et m_2

1.4.3 La règle de combinaison de Dempster

La règle de combinaison de Dempster permet de modifier les croyances après l'arrivée d'une nouvelle croyance. Cette règle permet de calculer, à partir de plusieurs fonctions de crédibilité sur le même ensemble de discernement Ω , une nouvelle fonction de crédibilité dite somme orthogonale en combinant les croyances.

D'une manière formelle, soient m_1, m_2 deux distributions de masses définies sur un référentiel Ω . Soient A_1, A_2, \dots, A_n les éléments focaux de m_1 et B_1, B_2, \dots, B_l les éléments focaux de m_2 . La distribution de masse $m = m_1 \oplus m_2$ est définie par :

$$m(\emptyset) = 0$$

$$\forall A \neq \emptyset, m(A) = \frac{1}{k} \sum_{A_i \cap B_j = A} m_1(A_i) * m_2(B_j) \quad (1.12)$$

où

$$k = 1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) * m_2(B_j)$$

k est une constante de normalisation et la quantité :

$$\sum_{A_i \cap B_j = \emptyset} m_1(A_i) * m_2(B_j)$$

mesure le degré de conflits entre les deux distributions de masse m_1 et m_2 .

La figure 1.1 [6] illustre la combinaison orthogonale de deux masses m_1 et m_2 .

Si $\sum_{A_i \cap B_j = \emptyset} m_1(A_i) * m_2(B_j) = 1$ alors les deux distributions de masses sont contradictoires et la combinaison est impossible.

La masse associée à un sous-ensemble P de Ω est donnée par :

$$m(P) = \sum_{A_i \cap B_j = P} m_1(A_i) * m_2(B_j)$$

1.4.4 Le conditionnement

Différentes définitions ont été développées pour le conditionnement dans la théorie des fonctions de croyance. Parmi lesquelles nous trouvons la règle de Dempster :

$$Pl(A | B) = \frac{Pl(A)}{Pl(B)} \quad (1.13)$$

1.4.5 Les fonctions de croyances et les probabilités

Nous pouvons faire le parallèle entre fonction de croyance et probabilités en plusieurs points :

- Si tous les éléments focaux associés à une croyance sont des singletons alors la mesure de probabilité est à la fois une mesure de croyance et une mesure de plausibilité.
- L'axiome d'additivité des probabilités est remplacé par l'inégalité suivante :

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B)$$

- La théorie des fonctions de croyance permet de distinguer l'ignorance de l'incertitude, contrairement à la théorie des probabilités. En effet, pour la théorie des fonctions de croyance, l'ignorance est représentée par le fait que deux événements peuvent être tous les deux plausibles sans être pour autant crédibles. Plus formellement, nous pouvons avoir :

$$Pl(A) = Pl(\neg A) = 1$$

et

$$Bel(A) = Bel(\neg A) = 0$$

Ainsi, la théorie des fonctions de croyance permet de distinguer entre l'ignorance et l'incertitude, contrairement à la théorie des probabilités.

- La théorie des probabilités est, sous certaines conditions, un cas particulier de la théorie des fonctions de croyance. En effet, les mesures de croyance et de plausibilité peuvent être vues respectivement comme étant la borne inférieure et la borne supérieure de la probabilité [21] :

$$\forall A \in \Omega, Bel(A) \leq P(A) \leq Pl(A)$$

Ainsi, l'intervalle $[Bel(A), Pl(A)]$ est interprété comme un encadrement de la probabilité réelle de A [17].

- Dans le cas où la distribution de masse m est définie sur les singletons du référentiel Ω et non pas sur les sous ensembles de Ω , elle devient équivalente à une distribution de probabilités. Ceci induit une égalité entre la mesure de probabilités, la mesure de crédibilité et la mesure de plausibilité. Ainsi,

$$\forall A \in \Omega, Bel(\{A\}) = Pl(\{A\}) = P(A) \quad (1.14)$$

L'équation de la théorie de Dempster-Shafer

$$Pl(A) = 1 - Bel(\neg A)$$

devient alors

$$P(A) + P(\neg A) = 1$$

C'est la propriété caractéristique de la théorie des probabilités.

Les différentes relations entre les mesures relatives à la théorie des croyances peuvent être illustrées sous forme graphique appelée intervalle de croyance (voir la figure 1.2). Nous pouvons souligner l'existence des fonctions symétriques de la crédibilité et de la plausibilité à savoir l'incrédulité (crédibilité de la négation) et le doute (la plausibilité de la négation).

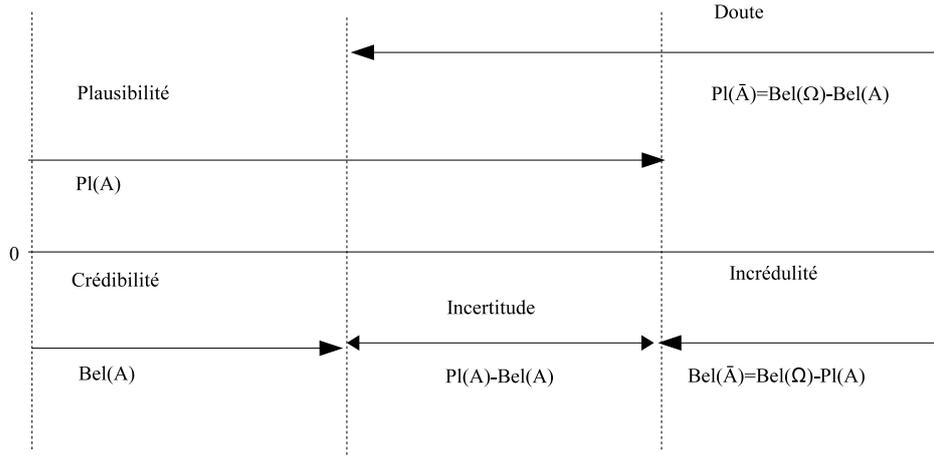


FIGURE 1.2 – Intervalle de croyance d'un élément focal A

1.4.6 Conclusion sur les fonctions de croyances

Les fonctions de croyances permettent la modélisation des connaissances incertaines incluant le caractère de "non monotonie" du raisonnement où une connaissance peut être revue après l'avènement d'une croyance grâce à la règle de conditionnement. Cette théorie fût populaire [22] et le modèle de base a été étendu dans de nombreuses directions [23, 24]. Entre autre, dans [25], l'auteur a développé une théorie de l'incertain non probabiliste appelé le modèle de croyance transférable qui est une interprétation particulière de la théorie des fonctions de croyance où la représentation s'effectue sur la base des informations partielles. Dans cette approche, le concept d'additivité est remplacé par l'inégalité suivante :

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B)$$

Cependant, la théorie des fonctions de croyances ne permet pas de prendre en compte les informations contradictoires.

1.5 La Théorie des possibilités

Une manière d'évaluer le degré de confiance attribué à un événement incertain était de déterminer le degré de certitude de l'événement contraire comme pour la théorie des probabilités. Une autre approche consiste à évaluer l'incertitude accordée à un événement par deux mesures distinctes, comme le cas de la théorie des possibilités.

La théorie des possibilités proposée par Zadeh [8] puis développée par Dubois et Prade [9] est une théorie de l'incertain, ayant pour vocation de manipuler des connaissances incomplètes [26] décrites en termes d'ensembles flous. Elle diffère de la théorie des probabilités vu qu'elle manipule deux mesures duales (mesure de possibilités et mesure de

nécessité). De plus, c'est une théorie non additive permettant de prendre en compte la notion de structures ordinales. Cette théorie a été développée dans deux directions : qualitative et quantitative. Ces deux approches partagent l'axiome de *maxitivité* mais diffèrent dans leurs définitions des concepts de conditionnement et d'indépendance.

Les mesures de possibilités qualitatives permettent d'évaluer les connaissances en utilisant la notion d'ordre. La logique possibiliste est issue de la théorie des possibilités. C'est une extension de la logique classique offrant ainsi un contexte sémantique pour le raisonnement non monotone dont la complexité reste proche à celle de la logique classique. Les échelles ne sont pas numériques mais ordinales : les connaissances sont organisées en niveaux (strates) selon leurs degrés de certitude. Ainsi, les formules de même degré occupent la même strate.

La théorie des possibilités quantitative possède diverses branches : le degré de possibilité peut être vu comme la borne supérieure d'une probabilité. Une distribution de possibilités peut être considérée comme une fonction *likelihood* et peut codifier une distribution de probabilités avec des valeurs extrêmes [26].

La théorie des possibilités repose sur les concepts décrits dans les sous-sections suivantes :

1.5.1 Distribution de possibilités

La théorie des possibilités est basée sur la notion de distribution de possibilités qui représente une fonction de Ω vers $[0, 1]$. Une distribution de possibilités π représente un état sur les connaissances permettant de distinguer ce qui est plausible à partir des connaissances les moins plausibles.

La distribution de possibilités présente les propriétés suivantes :

- $\pi(\omega) > 0$: ω est quelque peu possible,
- $\pi(\omega) = 0$: ω est impossible,
- $\pi(\omega) = 1$: ω est totalement possible,
- $\pi(\omega) > \pi(\omega')$: ω est plus plausible que ω' .

π est dite normalisée s'il existe une interprétation ω telle que $\pi(\omega)=1$. La condition de normalisation de la distribution de possibilités reflète la cohérence des connaissances représentées par π . Dans ce document, les distributions de possibilités considérées sont normalisées.

Soient π et π' deux distributions de possibilités. Si $\forall \omega, \pi(\omega) < \pi'(\omega)$, π est dite plus spécifique que π' ou que π' est plus générale ou moins informative.

Si $\forall \omega \neq \omega_0, \pi(\omega)=0$ et $\pi(\omega_0) = 1$ (toutes les interprétations de Ω sont impossibles sauf une), alors la connaissance du monde est complète.

Ces distributions sont les plus spécifiques.

1.5.2 Les mesures de possibilités et de nécessité

La théorie des possibilités a été définie afin de représenter des connaissances incomplètes, imprécises et incertaines. Comme pour la théorie des probabilités, l'évaluation des degrés de confiance d'un événement est numérique. Le degré de possibilité est défini par :

$$\Pi(\phi) = \max\{\pi(\omega) : \omega \models \phi\} \quad (1.15)$$

$\Pi(\phi)$ permet d'évaluer la cohérence de ϕ avec les connaissances disponibles représentées par π .

Exemple 1.8 Dans le domaine des prévisions météorologiques, si le mois en cours est Août, alors il est totalement possible que le ciel soit dégagé vu qu'il n'y a pas d'informations qui contredisent ce fait (exemple prévisions d'orages).

Le degré de possibilité Π satisfait les propriétés suivantes :

- $\Pi(\perp)=0$,
- $\Pi(\top)=1$,
- $\Pi(\phi)=0$: ϕ est impossible,
- $\Pi(\phi)=1$: ϕ est complètement possible,
- $\Pi(\phi)=\Pi(\neg\phi)=1$: ϕ et $\neg\phi$ sont totalement possibles (cas de l'ignorance totale),
- $\max(\Pi(\phi),\Pi(\neg\phi))=1$: ϕ ou $\neg\phi$ doit être possible (critère de normalisation),
- $\forall\phi,\forall\psi \Pi(\phi\vee\psi)=\max(\Pi(\phi),\Pi(\psi))$.

Les relations qui lient $\Pi(\phi)$ et $\Pi(\neg\phi)$ induisent la prise en compte de ces deux quantificateurs pour la description de la croyance sur l'occurrence d'un événement. $\Pi(\neg\phi)$ informe sur la possibilité de $\neg\phi$, donc sur la certitude ou la nécessité de l'occurrence de $\neg\phi$ vu que si $\neg\phi$ est impossible alors ϕ est certaine. Cette dualité permet de définir le degré de nécessité comme suit :

$$N(\phi) = 1 - \Pi(\neg\phi) = \min\{1 - \pi(\omega), \omega \models \neg\phi\}$$

Cette relation duale entre Π et N exprime le fait que plus ϕ est certaine moins $\neg\phi$ devient cohérente avec les croyances disponibles.

Par exemple si un étudiant veut faire de la recherche en intelligence artificielle, il est certain qu'il a des bases solides en logique.

Le degré de nécessité présente les propriétés suivantes :

- $N(\phi)=0$: ϕ n'est pas certaine, néanmoins ϕ peut être possible,
- $N(\phi)=1$: ϕ est certainement vraie,
- $N(\phi)>0 \Rightarrow N(\neg\phi)=0$: si ϕ est quelque peu certaine alors son contraire n'est pas du tout certain,
- $\Pi(\phi) < 1 \Rightarrow N(\phi)=0$: si ϕ n'est pas complètement possible alors ϕ est complètement incertaine,
- $\forall\phi,\forall\psi N(\phi\wedge\psi)=\min(N(\phi),N(\psi))$.

Remarque 1.3 Le domaine de ces différentes mesures est fini.

1.5.3 Conditionnement possibiliste

Le conditionnement est un concept fondamental pour la mise à jour des pondérations initiales attribuées aux connaissances représentées par la distribution des possibilités après l'arrivée d'une information certaine, notée ϕ .

Soit π la distribution de possibilités initiales. Vu que la théorie des possibilités offre deux directions (qualitative et quantitative), le conditionnement possibiliste admet alors deux définitions selon le contexte ordinal ou numérique (pour plus de détails, voir [27, 28, 29]).

- Dans le contexte ordinal ou qualitatif, nous assignons au meilleur modèle de ϕ le degré maximal de possibilités comme suit :

$$\Pi(\omega \mid \phi) = \begin{cases} 1 & \text{si } \pi(\omega) = \Pi(\phi) \text{ et } \omega \models \phi \\ \pi(\omega) & \text{si } \pi(\omega) < \Pi(\phi) \text{ et } \omega \models \phi \\ 0 & \text{sinon} \end{cases} \quad (1.16)$$

Cette définition correspond au conditionnement basé sur le minimum.

- Dans le contexte numérique, tous les éléments de ϕ sont augmentés proportionnellement :

$$\Pi(\omega \mid \phi) = \begin{cases} \frac{\pi(\omega)}{\Pi(\phi)} & \text{si } \omega \models \phi \\ 0 & \text{sinon} \end{cases} \quad (1.17)$$

Cette définition correspond au conditionnement basé sur le produit. Les deux définitions du conditionnement satisfont l'équation suivante qui rappelle l'équation de Bayes :

$$\forall \omega, \pi(\omega) = \pi(\omega \mid \phi) \otimes \Pi(\phi)$$

où \otimes représente le minimum (équation 1.16) ou le produit (équation 1.17).

Exemple 1.9 *Considérons deux variables binaires, relatives à la dégradation de la couche d'ozone (DO) et la pollution industrielle (PI) telles que : $D_{Do} = \text{Oui}(o), \text{Non}(\neg o)$ et $D_{Pi} = \text{Faible}(f), \text{Élevée}(\neg f)$. La table 1.1 donne les distributions de possibilités initiales. Supposons qu'une nouvelle information certaine arrive relative au fait que la pollution soit élevée. La croyance est représentée par ϕ . $\Pi(\phi) = \max(0.2, 0.5) = 0.5$.*

TABLE 1.1 – Distributions initiales

Do	Pi	$\Pi(\text{Do} \wedge \text{Pi})$
<i>o</i>	<i>f</i>	0.6
<i>o</i>	$\neg f$	0.2
$\neg o$	<i>f</i>	1
$\neg o$	$\neg f$	0.5

En utilisant l'équation du conditionnement basé sur le minimum, nous obtenons les distributions données par la table 1.2. En utilisant l'équation du conditionnement basé sur

TABLE 1.2 – Conditionnement basé sur le minimum

Do	Pi	$\Pi(\text{Do} \wedge \text{Pi} \mid \phi)$
<i>o</i>	<i>f</i>	0
<i>o</i>	$\neg f$	0.2
$\neg o$	<i>f</i>	0
$\neg o$	$\neg f$	1

le produit, nous obtenons les distributions données par la table 1.3.

TABLE 1.3 – Conditionnement basé sur le produit

Do	Pi	$\Pi(\text{Do} \wedge \text{Pi} \mid \phi)$
<i>o</i>	<i>f</i>	0
<i>o</i>	$\neg f$	0.4
$\neg o$	<i>f</i>	0
$\neg o$	$\neg f$	1

1.5.4 Mesures de probabilités et Mesures de possibilités

La théorie des possibilités est utilisée pour le raisonnement sous incertitude basé sur le concept d'ordre (qualitatif) aussi bien que sur le concept numérique (quantitatif), contrairement à la théorie des probabilités qui est essentiellement numérique.

Dans le contexte subjectif [3], la théorie des possibilités quantitative concurrence la théorie des probabilités. Le terme subjectif signifie que les probabilités et les théories numériques proposées pour représenter l'incertitude sont considérées comme des outils permettant de quantifier les croyances d'un agent sans tenir compte de la nature aléatoire ou fréquentiste des événements.

La table 1.4 présente une comparaison entre les deux théories par rapport à leurs concepts respectifs.

TABLE 1.4 – Comparaison entre la théorie des probabilités et la théorie des possibilités

Probabilités	Possibilités
Probabilité P Distribution p	Mesures Π et N Distribution π
Normalisation : $\sum_{\omega \in \Omega} P(\omega) = 1$	Normalisation : $\max_{\omega \in \Omega} \pi(\omega) = 1$
Marginalisation : $P(A) = \sum_{\omega} p(\omega)$	Marginalisation $\Pi(A) = \sup_{\omega} \pi(\omega)$
$P(A \cup B) = P(A) + P(B)$ si $A \cap B = \emptyset$	$\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$
$P(A \cap B) = P(A) * P(B)$ si A et B sont indépendants	$N(A \cap B) = \min(N(A), N(B))$
$P(A) + P(\bar{A}) = 1$	$\max(\Pi(A), \Pi(\bar{A})) = 1$ $\Pi(A) + \Pi(\bar{A}) \geq 1$ $N(A) + N(\bar{A}) \leq 1$
$\forall \omega \in \Omega, p(\omega) = \frac{1}{ \Omega }$	Ignorance : $\forall \omega \in \Omega, \pi(\omega) = 1$

1.5.5 Mesures de possibilités et les fonctions de croyance

Dans le cas où tous les éléments focaux sont emboîtés, c'est à dire si :

$$A_1 \subseteq A_2 \dots \subseteq A_n$$

alors, les mesures de croyance et de plausibilité deviennent équivalentes aux mesures de nécessité et de possibilité par les relations suivantes :

$$\forall A \subset \Omega, \forall B \subset \Omega, Bel(A \cap B) = \min(Bel(A), Bel(B)) \quad (1.18)$$

$$\forall A \subset \Omega, \forall B \subset \Omega, Pl(A \cup B) = \max(Pl(A), Pl(B)) \quad (1.19)$$

Une famille d'ensembles est dite emboîtée lorsque ses éléments sont ordonnés de telle sorte que chacun soit contenu dans le suivant. En pratique, dans une approche système expert, cette condition signifie que les informations fournies par les experts sont naturellement ordonnées au sens de leur précision car elles correspondent à des ensembles emboîtés de valeurs : $A_1 \subseteq A_2 \subseteq \dots \subseteq A_k$ [16]. Les différentes relations qui existent entre ces différentes mesures de l'incertitude sont illustrées par la figure 1.3.

1.5.6 Conclusion

La théorie des possibilités permet une représentation naturelle de la notion d'incertitude. L'incertitude associée à un événement est évaluée par deux mesures duales (le degré de nécessité et le degré de possibilités). De plus, elle offre deux contextes de travail, l'un quantitatif ou numérique et l'autre qualitatif ou ordinal. Ainsi, le choix du mode dépendra, par exemple, de la nature des connaissances incertaines à modéliser.

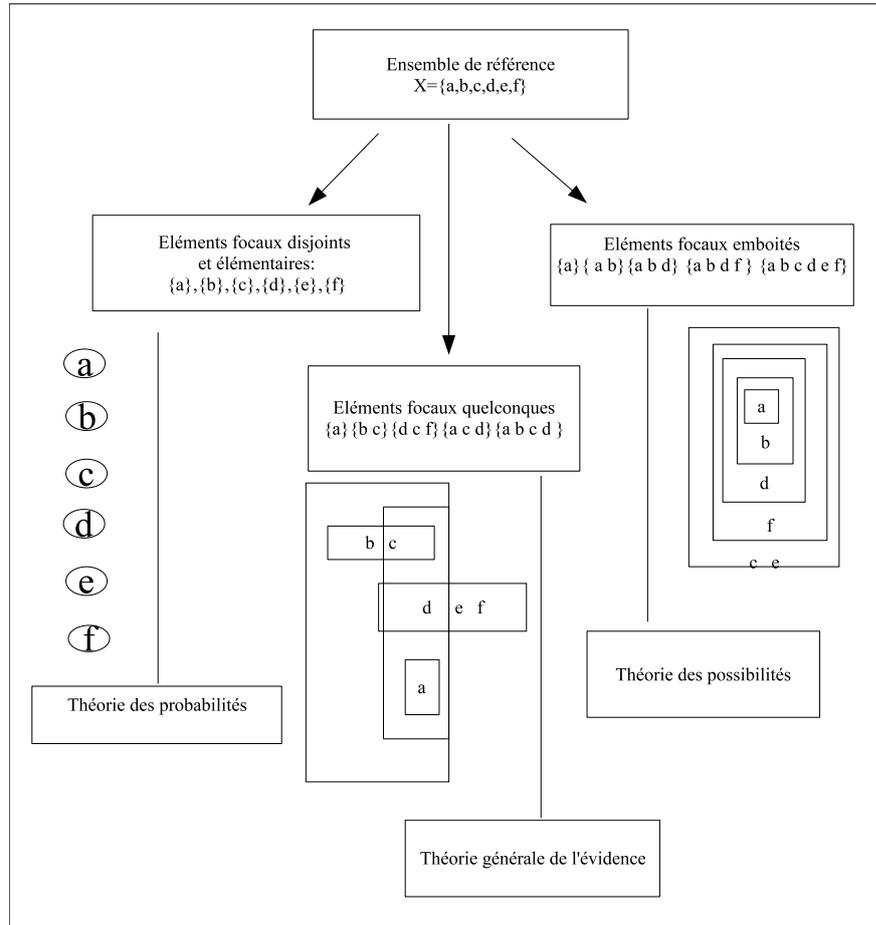


FIGURE 1.3 – Relations entre les modèles de l'incertitude illustrées par un exemple

1.6 Théorie des ensembles flous

La notion d'ensemble flou provient du constat établi par Zadeh que "très souvent, les classes d'objets rencontrés dans le monde physique ne possèdent pas de critères d'appartenance bien définis". La définition des ensembles flous répond au besoin de représenter des connaissances imprécises, floues et vagues qu'offre les langues naturelles du type "la plupart des suédois mesure plus de 1.75 m", "Mohammed est de grande taille". Dans cette proposition, le qualificatif grand est un prédicat imprécis vu qu'il n'y a pas de transitions marquées entre l'appartenance à un ensemble "grandes tailles" et la non appartenance. De tels ensembles sont dits flous.

Les principales notions de base de cette théorie [7] sont illustrées dans les prochaines sous-sections.

1.6.1 Fonctions d'appartenance

Afin de représenter des concepts flous, Zadeh [7] a défini la fonction caractéristique μ comme suit :

Définition 1.7 *Un ensemble flou A sur un référentiel Ω est caractérisé par une fonction d'appartenance μ_A qui associe à chaque élément x de Ω un nombre réel dans l'intervalle*

$[0,1]$, c'est à dire :

$$\mu_A : \Omega \rightarrow [0, 1]$$

La valeur $\mu_A(\omega)$ représente le degré d'appartenance de ω à A. Ainsi, plus la valeur $\mu_A(\omega)$ est proche de 1, plus grand est le degré d'appartenance de ω à A. En particulier :

- $\mu_A(\omega)=0$ représente la non-appartenance absolue,
- $\mu_A(\omega)=1$ représente l'appartenance totale,
- $0 < \mu_A(\omega) < 1$ représente l'appartenance nuancée ou partielle.

Un ensemble flou est dit normalisé s'il existe au moins un élément de son référentiel qui a son degré d'appartenance égale à 1. Il est à noter que la théorie des possibilités fait appel à des ensembles flous normalisés.

Nous illustrons cette notion d'ensemble flou par l'exemple suivant :

Exemple 1.10 Soit A l'ensemble flou des tailles sur le référentiel des tailles illustré par la figure 1.4. Elle indique :

- qu'une personne telle que sa taille est inférieure à 1,6 m est de petite taille,
- qu'une personne est d'autant plus grande que la taille se rapproche de 1.90 m,
- qu'une personne est considérée très grande si sa taille dépasse les 1.90 m.

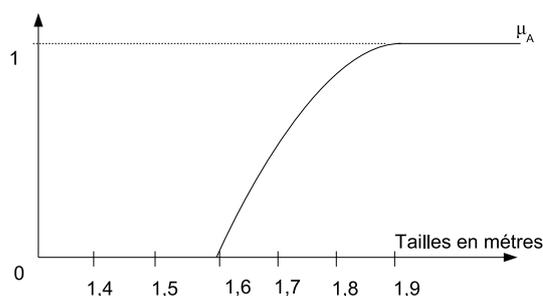


FIGURE 1.4 – Exemple d'ensemble flou "grandes tailles"

Remarque 1.4 La fonction d'appartenance n'a pas un caractère universel, mais dépend du contexte. Si nous prenons l'exemple de l'ensemble flou "grand", il diffère lorsque nous parlons de personnes, d'animaux ou autres.

1.6.2 Les caractéristiques d'un ensemble flou

Les caractéristiques d'un ensemble flou A sont les suivantes[5].

1. La première caractéristique est le support de A, noté $\text{Supp}(A)$ qui représente l'ensemble des éléments ω qui ont un degré d'appartenance positif. Plus formellement, le support de A est défini par :

$$\text{Supp}(A) = \{\omega \in \Omega : \mu_A(\omega) > 0\}$$

2. La deuxième caractéristique de A est sa hauteur notée $h(A)$ qui représente le plus fort degré avec lequel un élément de X appartient à A. Plus formellement, la hauteur de A est définie par :

$$h(A) = \{\sup_{\omega \in \Omega} \mu_A(\omega)\}$$

Ainsi, un ensemble flou est dit normalisé si sa hauteur est égale à 1.

3. Le noyau de A, noté $\text{Noy}(A)$ est l'ensemble des éléments de Ω pour lesquels la fonction d'appartenance de A vaut 1. Il est défini par :

$$\{\omega \in \Omega : \mu_A(\omega) = 1\}$$

4. Lorsque Ω est fini, l'ensemble flou A de Ω est caractérisé par sa cardinalité, qui indique le degré global avec lequel les éléments de Ω appartiennent à A. Elle est définie par :

$$|A| = \left\{ \sum_{\omega \in \Omega} \mu_A(\omega) \right\}$$

1.6.3 Les coupes de niveau

La représentation d'ensemble flou par une fonction d'appartenance peut être vue comme "une représentation verticale" des ensembles flous. Une autre vue possible, est celle qui consiste à considérer un ensemble flou comme une famille de sous-ensembles classiques emboîtés, à l'aide de la notion de *coupe de niveau* [30]. Pour $\alpha \in]0,1]$, une coupe de niveau α d'un ensemble flou A est définie par :

$$A_\alpha = \{\omega \in \Omega : \mu_A(\omega) \geq \alpha\}$$

Nous pouvons aussi définir une α -coupe stricte par :

$$A_\alpha^- = \{\omega \in \Omega : \mu_A(\omega) > \alpha\}$$

1.6.4 Les opérations ensemblistes

Zadeh a défini [7] des opérations ensemblistes sur les ensembles flous qui sont en fait une extension des opérations sur les ensembles classiques :

1.6.4.1 Égalité et inclusion des ensembles flous

- Deux ensembles flous A et B sont égaux si leurs fonctions d'appartenance prennent les mêmes valeurs pour tout élément.

$$\forall \omega \in \Omega, \mu_A(\omega) = \mu_B(\omega) \quad (1.20)$$

- Un ensemble flou A est dit inclus dans un ensemble flou B si leurs fonctions d'appartenance sont telles que :

$$\forall \omega \in \Omega, \mu_A(\omega) \leq \mu_B(\omega) \quad (1.21)$$

1.6.4.2 Intersection et Union des ensembles flous

- L'intersection de deux ensembles flous A et B est l'ensemble flou C noté $A \cap B$, tel que

$$\forall \omega \in \Omega, \mu_C(\omega) = \min(\mu_A(\omega), \mu_B(\omega)) \quad (1.22)$$

- L'union de deux sous ensembles flous A et B est l'ensemble flou D noté $A \cup B$, tel que :

$$\mu_D(\omega) = \max(\mu_A(\omega), \mu_B(\omega)) \quad (1.23)$$

TABLE 1.5 – Exemples de T-normes

T-normes	Nom
$T_1(a,b)=\min(a,b)$	Minimum
$T_2(a,b)=a*b$	Produit
$T_3(a,b)=\max(a+b-1,0)$	T-norme de Lukasiewicz

TABLE 1.6 – Exemples de Co-normes

Co-normes	Nom
$S_1(a,b)=\max(a,b)$	Maximum
$S_2(a,b)=a+b-a*b$	Somme probabiliste
$S_3(a,b)=\min(1,a+b)$	Somme bornée

D'une manière générale, l'intersection et l'union de deux ensembles flous peuvent être définies comme suit :

$$\mu_{A \cap B}(\omega) = T(\mu_A(\omega), \mu_B(\omega))$$

$$\mu_{A \cup B}(\omega) = S(\mu_A(\omega), \mu_B(\omega))$$

où T et S désignent respectivement une T-norme et une Co-norme. Les T-normes et les Co-normes les plus utilisées sont résumées dans les tables suivantes :

Les différentes T-normes et Co-normes vérifient les relations suivantes :

$$T_3 < T_2 < T_1$$

$$S_1 < S_2 < S_3$$

En pratique, les opérateurs min et max sont les plus utilisés pour modéliser respectivement l'intersection et l'union car ils vérifient la propriété d'idempotence.

1.6.4.3 Complément d'un ensemble flou

Le complément \bar{A} d'un ensemble flou A est défini comme l'ensemble flou ayant comme fonction d'appartenance :

$$\mu_{\bar{A}}(\omega) = 1 - \mu_A(\omega) \quad (1.24)$$

Propriété 1.3 *Comme dans la théorie des ensembles classiques, les opérations citées précédemment vérifient les propriétés suivantes :*

- Les opérations d'union \cup et d'intersection \cap sont commutatives et associatives.
- Les opérations d'intersection et d'union vérifient les lois de De Morgan et de distributivité :

Les lois de De Morgan

$$(A \bar{\cup} B) = \bar{A} \cap \bar{B}$$

$$(A \bar{\cap} B) = \bar{A} \cup \bar{B}$$

Les lois de distributivité

$$C \cap (A \cup B) = (C \cap A) \cup (C \cap B)$$

$$C \cup (A \cap B) = (C \cup A) \cap (C \cup B)$$

- L'inclusion est réflexive et transitive

1.6.5 Mesures de possibilité et de nécessité

La théorie des possibilités est fondée sur la théorie des ensembles flous [8]. Elle se base sur la notion de distributions des possibilités et sur les mesures duales de possibilité et de nécessité.

Du point de vue mathématique, les mesures de possibilités et de nécessité se réfèrent aux ensembles flous [7, 8].

Distribution de possibilités Une distribution de possibilité, notée π_ω sur un référentiel U peut être vue comme une fonction caractéristique généralisée d'ensembles flous représentant les valeurs connues comme des valeurs plus ou moins possibles pour la variable considérée [8].

Définition 1.8 [8] Soit A un sous ensemble flou de l'univers du discours Ω caractérisé par sa fonction d'appartenance μ_A . Soit X une variable prenant ses valeurs sur Ω . La proposition " X est A " associée à X la distribution de possibilités π_x définie comme étant numériquement égale à la fonction d'appartenance de A :

$$\forall u \in \Omega, \pi_x(u) = \mu_A(u)$$

Cette égalité numérique n'implique pas l'égalité des deux concepts. En effet, $\pi_x(u)$ évalue la possibilité que $x=u$ étant donné l'état des connaissances incomplet " $x=a$ " alors que $\mu_A(u)$ évalue le degré de compatibilité de l'information $x=u$ avec la déclaration " x est A " [30].

Exemple 1.11 Nous savons de façon certaine que Jean a aux alentours de 30 à 33 ans et qu'en tout état de cause il n'a pas moins de 27 ans ni plus de 38 ans. Cette information imprécise peut être représentée par une distribution de possibilités $\mu_{age(Jean)}$ comme le montre la figure 1.5.

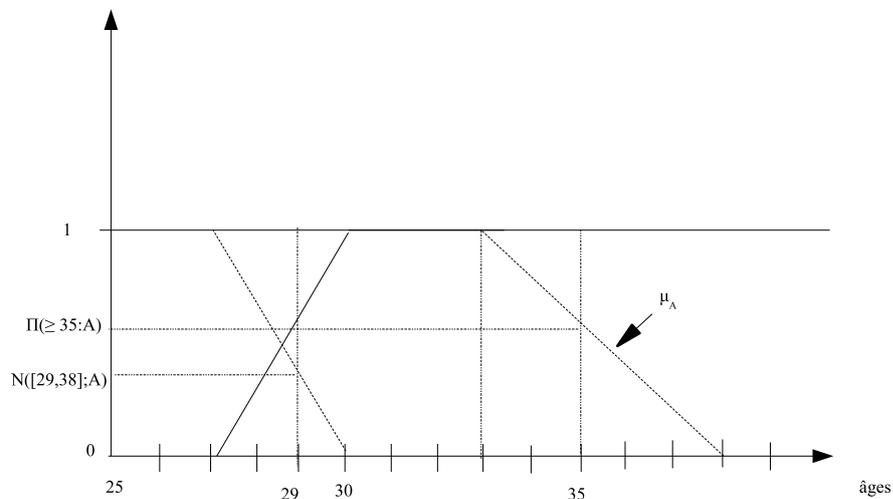


FIGURE 1.5 – Distribution de possibilités $\mu_{age(Jean)}$

Exemple 1.12 [3] $\mu_{Tall}(h)$ quantifie l'appartenance d'une personne de taille h à l'ensemble $Tall$, $\pi_{Tall}(h)$ quantifie la possibilité que la taille d'une personne est h avec l'hypothèse qu'elle appartient à l'ensemble $Tall$. Le principe possibiliste est traduit par l'égalité suivante :

$$\pi_{Tall}(h) = \mu_{Tall}(h)$$

avec $h \in H$ et H : ensemble des tailles $[0, +\infty]$

Elle stipule que la possibilité qu'un homme grand a la taille h est numériquement égale au degré d'appartenance d'un homme d'une taille h à l'ensemble des hommes grands.

$$\forall h \in H, \mu(Tall | h) = x \rightarrow \pi(h | Tall) = x, \quad (1.25)$$

Mesure de possibilité Zadeh a proposé de définir la possibilité d'un événement flou de la manière suivante [8] :

Définition 1.9 Soit A un ensemble flou de Ω . Soit π_x une fonction de distribution de possibilité associée à la variable X qui prend ses valeurs sur Ω . La mesure de possibilité, $\Pi(A)$ est définie par :

$$\Pi(A) = \sup_{u \in \Omega} \min(\mu_A(u), \pi_X(u))$$

où μ_A est la fonction d'appartenance de A .

Si A est non flou, nous retrouvons la définition de la possibilité d'un ensemble non flou. La mesure de possibilité peut être vue comme le degré (la hauteur) d'intersection de l'ensemble flou A et l'ensemble flou F_x défini par la fonction de distribution de possibilités π_x (voir figure 1.6).

$$\Pi(A) = 0 \Leftrightarrow \text{supp}(A) \cap \text{supp}(F_x) = \emptyset$$

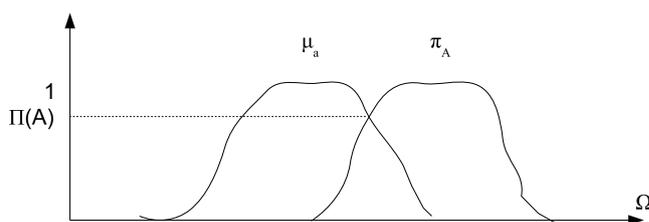


FIGURE 1.6 – Mesure de possibilité

1.6.6 Raisonnement approché

En logique classique, le modus ponens constitue le mécanisme de base de l'inférence déductive. Il stipule :

Soit R la règle "Si X est A alors Y est B " et F le fait " X est A ". La conclusion " Y est B " est alors inférée.

En présence de connaissances imprécises, cette règle n'est pas suffisante vu que les prémisses peuvent être juste partiellement vérifiées. Pour pallier à ce problème, Zadeh a défini le modus ponens généralisé [8]. C'est une règle d'inférence compositionnelle permettant de déduire une conclusion imprécise à partir de prémisses imprécises et de

faits imprécis.

Le modus ponens est représenté par le syllogisme suivant :

Règle : Si X est A alors Y est B

Fait : X est A'

Conclusion : Y est B'

Nous allons décrire dans ce qui suit le modus ponens généralisé qui fait appel aux notions d'implication ainsi qu'au mécanisme de combinaison/projection.

1.6.6.1 L'implication floue

L'implication floue entre deux propositions floues "x est A" et "y est B" est une proposition floue concernant le couple de variable (x,y) dont la valeur de vérité est donnée par la fonction d'appartenance μ_R d'une relation floue entre x et y définie par :

$$\forall (u, v) \in U \times V \mu_R(u, v) = I(\mu_A(u), \mu_B(v)) \quad (1.26)$$

où I est une fonction de $[0,1] \times [0,1] \rightarrow [0,1]$ qui désigne une implication floue et vérifie les propriétés suivantes :

- I_1 : $I(0,0)=I(0,1)=I(1,1)=1$; $I(1,0)=0$,
- I_2 : si $x \leq y$ alors $I(x,y) \geq I(z,y) \forall y \in [0,1]$,
- I_3 : si $y \leq t$ alors $I(x,y) \leq I(x,t) \forall x \in [0,1]$,
- I_4 : $I(0,x)=1 \forall x \in [0,1]$,
- I_5 : $I(x,1)=1 \forall x \in [0,1]$.

Différents types d'implications floues ont été proposées.

La table 1.7 regroupe les implications floues les plus utilisées.

TABLE 1.7 – Principales implications floues

Implication floue	Nom
$I(a,b)=\max(1-a,b)$	Kleene-Dienes
$I(a,b)=1-a+a*b$	Reichenbach
$I(a,b)=\min(1,1-a+b)$	Lukasiewicz
$I(a,b)=1$ Si $a \leq b$ $I(a,b)=b$ sinon	Gödel
$I(a,b)=1$ si $a \leq b$ $I(a,b)=b/a$ sinon	Goguen

1.6.6.2 Le modus ponens généralisé

Considérons une règle floue de la forme "Si X est A alors Y est B" et un fait "X est A'" avec A et A' deux ensembles flous de U et B un ensemble flou de V. Le processus de raisonnement approché consiste à déterminer une conclusion de la forme "Y est B'" à partir de ces deux informations, c'est à dire de quelle façon la différence entre A et A' doit être répercutée sur B pour obtenir B' [30]. La méthode proposée par Zadeh consiste à définir la relation R correspondant à la règle, à la combiner avec A' puis de projeter le résultat sur V. Le calcul de B' s'effectue par le processus de combinaison/projection. La première opération consiste à combiner les différentes distributions de possibilités π_i définies sur U_i associées aux différentes variables comme suit :

$$\pi = \min_{i=1,r} \pi_i$$

La projection de π sur le domaine U_i de x_i est définie de la manière suivante :

$$\forall u_k \in U_k, \pi_i(u_i) = \sup_{u_k, k \rightarrow i} \pi_{x_1, \dots, x_n}(U_1, \dots, U_n)$$

Ainsi, en appliquant le processus de combinaison/projection, nous obtenons :

$$\forall v \in V, \mu_{B'}(v) = \sup_{u \in U} \min(\mu_{A'}(u), \mu_R(u, v))$$

où R est une implication floue I ($\mu_R(u, v) = I(\mu_A(u), \mu_B(v))$).

1.6.6.3 Le contrôle flou

c'est l'une des applications pratiques de la logique floue. Un contrôleur flou se comporte comme un approximateur capable de calculer un résultat non flou depuis plusieurs variables non floues mais par le biais d'un processus de raisonnement flou. Il comporte trois phases :

- La fuzzification : consiste à transformer les données non floues en données floues.
- le noyau composé d'une base de connaissances floues et d'un moteur d'inférences floues.
- La defuzzification : consiste à transformer les sorties floues en des sorties non floues.

1.7 Conclusion

Les différents modèles mathématiques de l'incertitude présentés dans ce chapitre à savoir la théorie des probabilités, la théorie des croyances et la théorie des possibilités, sont les modèles de l'incertain les plus couramment utilisés en intelligence artificielle. Ils ont divers mérites dans le domaine de la représentation des connaissances incertaines. En utilisant la théorie des probabilités, l'ignorance totale sur un événement A ($P(A)=0$) entraîne que $\neg A$ doit être considéré comme certain ($P(\neg A)=1$) alors que dans les deux autres modèles, le fait que A ne soit pas vrai n'entraîne pas la croyance totale en l'événement $\neg A$.

Sous un autre angle, ces différentes facettes de l'incertitude sont liées. La figure 1.7 illustre la hiérarchie qui existe entre elles. Nous relevons plusieurs points :

1. La théorie des fonctions de croyances de Dempster-Shafer est un cadre général du raisonnement dans l'incertain qui permet de généraliser, entre autres, les deux cas particuliers que sont la théorie des possibilités et la théorie des probabilités [31] (voir figure 1.8).
2. Les degrés de confiance qui sont à la fois des mesures de possibilités, des mesures de crédibilité, de plausibilité et de probabilités sont des mesures de Dirac : $\exists \omega, \omega \in \Omega$ telle que $d(\omega) = 1$.
3. Toutes ces mesures sont elles mêmes des cas particuliers des mesures floues.

En intelligence artificielle, la logique a été considérée, depuis longtemps, à la base de la représentation et du raisonnement sur les connaissances d'une manière générale. Le chapitre suivant est destiné à exposer deux types de logiques quantitatives, qui nous intéressent particulièrement, modélisant les connaissances incertaines, à savoir la logique possibiliste quantitative et la logique des pénalités.

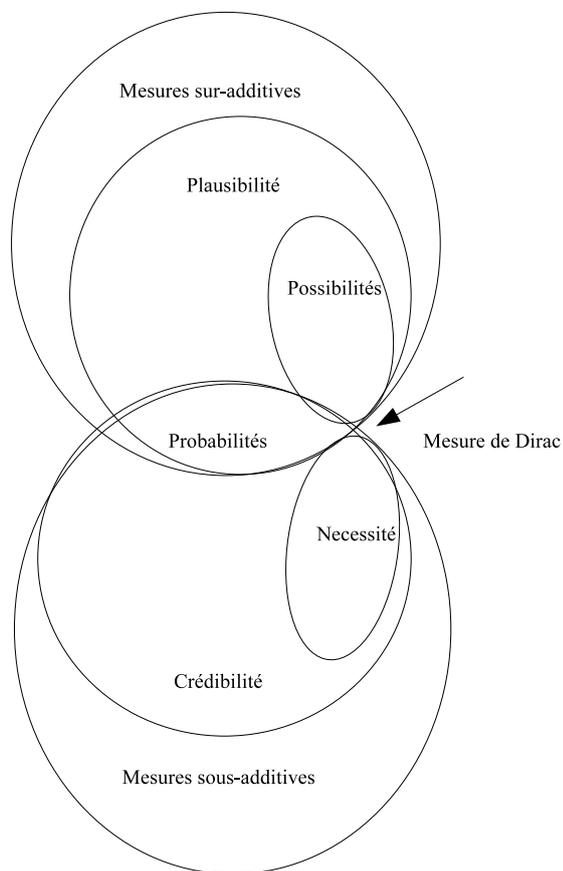


FIGURE 1.7 – Tableau récapitulatif sur les mesures de l'incertain

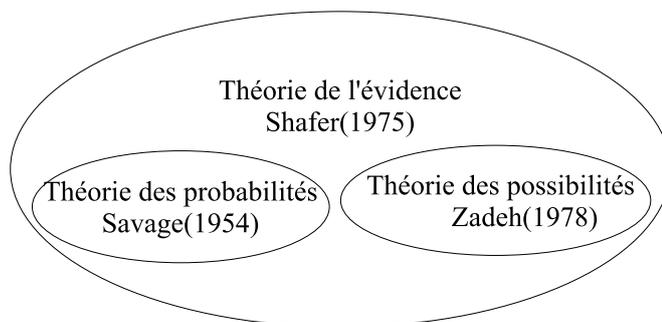


FIGURE 1.8 – Relation entre les modèles de l'incertain

Chapitre 2

Modèles Logiques de l'incertain

2.1 Introduction

Le développement des recherches dans le domaine de la logique, vu sa richesse d'expression permettant de spécifier des problèmes complexes, a toujours eu une incidence sur le domaine de l'intelligence artificielle en général et sur la représentation et le raisonnement dans un cadre incertain en particulier.

Dans le contexte de l'incertain, plusieurs logiques ont été développées. Parmi lesquelles nous trouvons :

- les logiques numériques qualitatives ou ordinales telles que la logique possibiliste standard,
- les logiques numériques quantitatives telles que la logique possibiliste quantitative et la logique des pénalités,
- les logiques non numériques comme les logiques non monotones, la logique des défauts, la logique de la circonscription,...etc.

Ce chapitre est consacré à donner un bref aperçu sur ces différentes logiques. La section 2 présente la logique possibiliste standard, tout en mentionnant les relations qui la lie à d'autres logiques. La section 3 est dédiée à la logique possibiliste quantitative. La section 4 expose les concepts de base de la logique des pénalités, tout en invoquant les relations qu'elle entretient avec d'autres approches.

2.2 La logique possibiliste qualitative

Comme nous l'avons vu dans le chapitre précédent, la théorie des possibilités offre deux contextes de travail, l'un ordinal et l'autre numérique. Ainsi, si un expert veut représenter ses connaissances en utilisant la logique, il dispose en fait de deux types de logique possibiliste : la logique possibiliste standard et la logique possibiliste quantitative.

La logique possibiliste standard est une logique de l'incertain conçue pour raisonner avec des connaissances incomplètes et partiellement inconsistantes [32]. Elle est basée sur l'idée d'ordre (complet) et non pas sur le calcul comme le cas des probabilités.

Au niveau syntaxique, des formules logiques propositionnelles sont manipulées, auxquelles sont attachées des valuations numériques correspondant au seuil minimal des degrés de nécessité.

Au niveau mathématique, les degrés de possibilité et de nécessité se réfèrent aux ensembles flous [7, 8] et la logique possibiliste est alors adaptée pour le raisonnement

automatique où les connaissances disponibles sont entachées de flou.

2.2.1 Base possibiliste standard

En logique possibiliste standard, les informations incertaines sont représentées par une base propositionnelle avec priorité, appelée base de croyances possibilistes [33]. Plus formellement, une base possibiliste Σ est constituée d'un ensemble de couples (ϕ_i, α_i) où ϕ_i est une formule propositionnelle et $\alpha_i \in [0,1]$ est le poids associé à chaque formule correspondant au seuil minimal du degré de nécessité. D'une manière formelle, $\Sigma = \{(\phi_i, \alpha_i) : i = 1, n\}$ avec $N(\phi_i) \geq \alpha_i$.

2.2.2 Distribution de possibilités associée à une base possibiliste standard

La base possibiliste est une représentation compacte des distributions de possibilités. Chaque élément d'information (ϕ_i, α_i) d'une base de connaissances peut être vu comme une contrainte qui restreint le degré de possibilité associé aux interprétations [34].

Si une interprétation ω satisfait ϕ_i alors son degré de possibilités est égal à 1 (ω est complètement compatible avec la croyance ϕ_i), sinon, le degré est égal à $1 - \alpha_i$ (plus ϕ_i est certaine, moins ω est possible). En particulier, si $\alpha_i = 1$, alors toute interprétation qui falsifie ϕ_i , est telle que son degré de possibilité est égal à 0, donc est impossible. D'une manière formelle, la distribution de possibilités associée à une formule pondérée (ϕ_i, α_i) est définie par [34] : $\forall \omega \in \Omega$:

$$\pi_{(\phi_i, \alpha_i)}(\omega) = \begin{cases} 1 - \alpha_i & \text{si } \omega \not\models \phi_i \\ 1 & \text{sinon} \end{cases} \quad (2.1)$$

Plus généralement, la distribution de possibilités associée à une base de connaissances possibiliste standard Σ est le résultat de la combinaison, par l'opérateur min, des distributions de possibilités associées à chaque formule pondérée (ϕ_i, α_i) de Σ , à savoir : $\forall \omega \in \Omega$:

$$\pi_{\Sigma}(\omega) = \begin{cases} 1 & \text{si } \forall (\phi_i, \alpha_i) \in \Sigma, \omega \models \phi_i \\ \min\{1 - \alpha_i : (\phi_i, \alpha_i) \in \Sigma, \omega \not\models \phi_i\} & \text{sinon} \end{cases} \quad (2.2)$$

Exemple 2.1 Soit la base de connaissances possibiliste standard suivante :

$\Sigma = \{(\neg a \vee b, .1), (a \vee b \vee c, .3), (\neg b \vee d, .5), (\neg d, .9)\}$. La distribution de possibilités est donnée dans la table 2.1.

2.3 La logique possibiliste quantitative

La logique possibiliste quantitative est une logique numérique. Ainsi, un degré d'incertitude, correspondant au seuil minimal du degré de nécessité est associé à chaque formule ϕ_i de la base de connaissances. Contrairement à la logique possibiliste standard, les quantifications numériques attachées aux formules n'induisent pas une relation d'ordre. En effet, la pondération α_i associée à chaque formule ϕ_i permet juste de quantifier son degré d'incertitude.

Soit $\Sigma = \{(\phi_i, \alpha_i) : i = 1, \dots, n\}$ avec $N(\phi_i) \geq \alpha_i$. La distribution de possibilités associée à une base de connaissances possibiliste quantitative Σ est définie par : $\forall \omega \in \Omega$:

$$\pi_{\Sigma}(\omega) = \begin{cases} 1 & \text{si } \forall (\phi_i, \alpha_i) \in \Sigma, \omega \models \phi_i \\ * \{1 - \alpha_i : (\phi_i, \alpha_i) \in \Sigma, \omega \not\models \phi_i\} & \text{sinon} \end{cases} \quad (2.3)$$

TABLE 2.1 – Distribution de possibilités associée à la base de connaissances de l'exemple 2.1

ω	$\pi_{\Sigma}(\omega)$
$a \wedge b \wedge c \wedge d$.1
$a \wedge b \wedge c \wedge \neg d$.5
$a \wedge b \wedge \neg c \wedge d$.1
$a \wedge b \wedge \neg c \wedge \neg d$.5
$a \wedge \neg b \wedge c \wedge d$.1
$a \wedge \neg b \wedge c \wedge \neg d$.5
$a \wedge \neg b \wedge \neg c \wedge d$.1
$a \wedge \neg b \wedge \neg c \wedge \neg d$.9
$\neg a \wedge b \wedge c \wedge d$.1
$\neg a \wedge b \wedge c \wedge \neg d$.5
$\neg a \wedge b \wedge \neg c \wedge d$.1
$\neg a \wedge b \wedge \neg c \wedge \neg d$.5
$\neg a \wedge \neg b \wedge c \wedge d$.1
$\neg a \wedge \neg b \wedge c \wedge \neg d$	1
$\neg a \wedge \neg b \wedge \neg c \wedge d$.1
$\neg a \wedge \neg b \wedge \neg c \wedge \neg d$.7

Exemple 2.2 Soit Σ la base de connaissances possibiliste quantitative suivante : $\Sigma = \{(a \vee b, .8), (\neg a \vee c, .2), (c \vee \neg d, .5)\}$

La distribution de possibilités associée à Σ est donnée par la table 2.2.

TABLE 2.2 – Distribution de possibilités associée à la base de connaissances de l'exemple 2.2

ω	$\pi_{\Sigma}(\omega)$
$a \wedge b \wedge \neg c \wedge d$.4
$a \wedge b \wedge \neg c \wedge \neg d$.8
$a \wedge \neg b \wedge \neg c \wedge d$.4
$a \wedge \neg b \wedge \neg c \wedge \neg d$.8
$\neg a \wedge b \wedge \neg c \wedge d$.5
$\neg a \wedge \neg b \wedge c \wedge d$.2
$\neg a \wedge \neg b \wedge c \wedge \neg d$.2
$\neg a \wedge \neg b \wedge \neg c \wedge d$.1
$\neg a \wedge \neg b \wedge \neg c \wedge \neg d$.2
Autres interprétations	1

Du côté du mode d'inférence, contrairement à la logique possibiliste standard, il n'y a pas eu de développement d'algorithmes de raisonnement. Pour ce faire, il faudrait tenir compte de la nature des connaissances exprimées dans le contexte de la logique possibiliste quantitative. Nous allons nous y intéresser dans le chapitre 8.

2.4 La logique des pénalités

La logique des pénalités a été proposée par Pinkas [35] et développée par Dupin de Saint-Cyr [36, 37, 31]. Dans cette approche, à chaque formule de la base de connaissances est associée une pénalité, qui correspond au prix à payer pour pouvoir rejeter cette

formule. Plus la pénalité est élevée, plus la formule est importante.

D'une manière plus formelle, une base de connaissances à pénalités est définie par :

Définition 2.1 Soit \mathcal{R}^{*+} l'ensemble des réels strictement positifs plus $\{+\infty\}$. Une base de connaissance à pénalités notée PK est un multi-ensembles fini de couples (ϕ_i, α_i) avec $\phi_i \in \mathcal{L}$ et $\alpha_i \in \mathcal{R}^{*+}$, α_i étant la pénalité associée à ϕ_i .

Dans une optique de raisonnement, α_i représente le degré de certitude associé à ϕ_i . Si $\alpha_i = +\infty$ alors ϕ_i est complètement certaine. Si une formule apparaît plusieurs fois dans une base, la pénalité résultante est la somme de l'ensemble des pénalités associées à la formule.

Exemple 2.3 $PK = \{(a, 1), (a, 1)\}$ n'est pas équivalente à $PK' = \{(a, 1)\}$ vu que dans la base PK le coût pour supprimer a est égale à 2 alors que dans la base PK' le coût est égal à 1.

Remarque 2.1 : Si dans une base à pénalités, toutes les pénalités sont infinies, cela implique qu'aucune formule ne peut être violée, ce qui réduit la logique des pénalités à la logique classique.

2.4.1 Coût d'une interprétation

Soit $PK = \{(\phi_i, \alpha_i) : i = 1, \dots, n\}$ une base de connaissances à pénalités.

Définition 2.2 [35]

Le coût d'une interprétation $\omega \in \Omega$ conformément à PK noté k_{PK} est égal à la somme des pénalités des formules de PK violées par ω :

$$k_{PK}(\omega) = \begin{cases} 0 & \text{Si } \forall (\phi_i, \alpha_i) \in PK, \omega \models \phi_i \\ \sum \{\alpha_i : (\phi_i, \alpha_i) \in PK, \omega \not\models \phi_i\} & \text{Sinon} \end{cases} \quad (2.4)$$

Exemple 2.4 Soit $PK = \{(\neg a, 6), (\neg b \vee c, 8), (a \vee c, 3), (\neg c, +\infty)\}$ Les différents coûts d'interprétation sont donnés par la table 2.3.

TABLE 2.3 – Coûts d'interprétation de l'exemple 2.4

ω	k_{PK}
$a \ b \ c$	$+\infty$
$a \ b \ \neg c$	$6+8=14$
$a \ \neg b \ c$	$+\infty$
$a \ \neg b \ \neg c$	6
$\neg a \ b \ c$	$+\infty$
$\neg a \ b \ \neg c$	$3+8=11$
$\neg a \ \neg b \ c$	$+\infty$
$\neg a \ \neg b \ \neg c$	3

Une base à pénalités induit une distribution de pénalités unique $k_{PK}(\omega)$. $k_{PK}(\omega)$ est d'autant plus faible qu'il est plausible que ω soit le monde réel.

2.4.2 Coût d'une formule

Le coût d'une formule ϕ cohérente, noté $K_{PK}(\phi)$ est le coût minimal relativement à PK qui satisfait ϕ :

$$K_{PK}(\phi) = \min\{k_{PK}(\omega) : \omega \models \phi\} \quad (2.5)$$

Le coût k_{PK} satisfait les propriétés suivantes :

Propriété 2.1 $K_{PK}(\perp) = +\infty$

Cette propriété est naturelle, vu qu'il n'est pas possible de rendre K_{PK} cohérent avec \perp .

Propriété 2.2 $K_{PK}(\top) = \min\{k_{PK}(\omega) : \omega \in \Omega\}$

$K_{PK}(\top)$ représente le coût minimal pour rendre K_{PK} cohérente.

Propriété 2.3 $K_{PK}(\top) = +\infty \Leftrightarrow \forall (\phi_i, \alpha_i) \in PK, \alpha_i = +\infty$ et $\{(\phi_i, \alpha_i) \in PK\}$ est incohérente.

Propriété 2.4 $K_{PK}(\top) = 0 \Leftrightarrow PK$ est cohérente.

Propriété 2.5 $\forall \phi, \psi \in \mathcal{L}, (\phi \models \psi) \rightarrow K_{PK}(\phi) \geq K_{PK}(\psi)$

Propriété 2.6 $\forall \phi, \psi \in \mathcal{L},$

1. $K_{PK}(\phi \wedge \psi) \geq \max(K_{PK}(\phi), K_{PK}(\psi))$
2. $K_{PK}(\phi \vee \psi) = \min(K_{PK}(\phi), K_{PK}(\psi))$
3. $K_{PK}(\perp) \geq K_{PK}(\phi) \geq K_{PK}(\top)$

Remarque 2.2 Dans [38], Freund a mis en exergue la relation qui existe entre la logique des pénalités et la logique des défauts. Il a proposé une approche additive pour générer une relation d'inférence non-monotone à partir d'un ensemble de règles de défauts. Cette approche associe d'abord une valuation entière à chaque règle. Ces valuations sont en fait des pénalités puisqu'elles induisent une valuation sur chaque interprétation ω en sommant les poids des règles violées par ω . Ainsi, il y a génération automatique des pénalités à partir d'un ensemble de défauts $\phi_i \rightarrow \psi_i$.

2.5 Les logiques connexes

Nous allons brièvement aborder dans cette section :

- les relations qui existent entre la logique possibiliste standard et certaines logiques numériques telles que la logique probabiliste et certaines logiques non numériques telles que la logique modale et la logique des défauts,
- les relations qu'entretient la logique des pénalités avec la logique possibiliste, les fonctions de croyance et les fonctions conditionnelles ordinales.

2.5.1 La logique possibiliste et la logique probabiliste

La logique probabiliste peut être utilisée pour modéliser le raisonnement automatique. En effet, le mécanisme d'inférence peut être étendu à la logique probabiliste. Ainsi par exemple le principe de résolution s'exprimerait par [28] :

$$P(q \vee r) \geq \max(0, P(p \vee q) + P(\neg p \vee r) - 1)$$

Il est à noter que la borne inférieure est plus petite si nous la comparons à celle utilisant les degrés de nécessité. De plus, l'utilisation de l'axiome d'additivité dans la logique probabiliste entraîne un taux d'erreur croissant des données de départ vers les conclusions, tandis que dans le contexte possibiliste, ce taux est constant. Le critère de normalisation dans la logique probabiliste ($\sum_{i=1..n} P(\omega_i)=1$ où ω_i sont des événements élémentaires) est plus difficile à satisfaire que la règle de normalisation pour les mesures de possibilités ($\max_{i=1..n} \Pi(\omega_i)=1$).

Néanmoins, il s'est avéré que les mesures de possibilités peuvent être interprétées comme une famille spéciale des probabilités inférieures et supérieures [27] et les fonctions ordinales conditionnelles [39] ont permis d'interpréter les mesures de possibilités en termes de probabilités infinitésimales [40].

La logique possibiliste et la logique probabiliste ne sont pas dédiées aux mêmes buts [34]. La logique probabiliste apparaît mieux adaptée pour traiter les connaissances incomplètes de type statistique. Elle a été également construite comme une théorie de l'induction [41] ou une logique pour les croyances subjectives [42]. La logique probabiliste est plus complexe que la logique possibiliste, spécialement si les probabilités conditionnelles doivent être prises en compte dans le langage. Ceci implique que le raisonnement possibiliste est plus simple. La logique possibiliste quant à elle a pour but de permettre de raisonner avec la partie la plus certaine de la base de connaissances.

2.5.2 La logique possibiliste et la logique modale

Il est également intéressant de discuter des liens qui existent entre la logique des possibilités et la logique modale qui procure une modélisation des modalités telles que la possibilité et la nécessité [28].

La logique modale a été introduite afin de prendre en compte des affaiblissements ou des renforcements d'affirmations présents dans les langues naturelles. Les principales différences entre ces deux approches sont :

- Dans la logique modale, possibilité et nécessité sont des concepts "tout ou rien". Ils peuvent être introduits comme des symboles spéciaux du langage. $\Box p$ correspondrait à "p est nécessaire" et $\Diamond p$ correspondrait à "p est possible". Dans la théorie des possibilités, les notions de possibilité et de nécessité sont des notions graduelles.
- La logique modale propose différents types d'axiomatisation, tandis que les axiomes de la théorie des possibilités sont bien définis et sont uniques. Dans cette optique, il est judicieux de définir la contrepartie qualitative des axiomes de la théorie des possibilités dans le style de la logique modale. Une façon de faire consiste à utiliser les règles de traductions suivantes :
 - $\vdash \Box P$ est traduit en $N(p)=1$
 - $\vdash \Diamond P$ est traduit en $\Pi(p)=1$
 - l'identité classique $\neg \Diamond p = \Box \neg p$ est traduite en $1-\Pi(p)=N(\neg p)$
- De plus, une traduction numérique de l'implication de Lewis $\Box(p \rightarrow q)$ serait $N(p \rightarrow q)=1$ qui implique que $p \rightarrow q$ est vrai dans la logique possibiliste.

Pour plus de détails sur les relations entre la logique possibiliste et la logique modale voir [43].

2.5.3 Raisonnement par défaut utilisant les logiques de l'incertain

La logique probabiliste et la logique possibiliste ont été suggérées comme des approches possibles pour le raisonnement par défaut [44, 45]. L'idée est d'interpréter le

poids accordé à une règle "Si-Alors" comme une mesure qui évaluerait à quel point cette règle n'a pas d'exception. Cette interprétation engendre plusieurs problèmes [28] :

- Les règles par défaut n'ont pas toujours une interprétation statistique. En particulier, la typicalité semble être de nature différente. Une règle par défaut est alors modélisée par la connaissance de la quantité $\Pi(q | p)$ définie par la relation :

$$\Pi(p \wedge q) = \min(\Pi(p), \Pi(q | p))$$

Néanmoins, cette notion de conditionnement est très liée à l'implication logique, vu que $\Pi(q | p) = 1 - N(p \rightarrow \neg q)$.

- Ce n'est pas clair que dans la logique des défauts, nous puissions attacher des degrés d'incertitude à une implication logique $p \rightarrow q$.

L'utilisation du conditionnement au lieu de l'implication peut apparaître une approche plus naturelle pour modéliser les relations imparfaites de type "est un".

2.5.4 La logique des pénalités et la logique possibiliste

Le coût d'une formule est une mesure de possibilité non normalisée avec un changement d'intervalle de $[0, +\infty]$ en $[0, 1]$. En effet [31], la logique des pénalités peut être considérée comme une logique multi-sources où chaque formule $\langle \phi_i, \alpha_i \rangle$ d'une base de connaissances à pénalité provient d'une source indépendante et exprime une contrainte sur une fonction coût k_i définie par :

$$k_i(\omega) = \begin{cases} \alpha_i & \text{si } \omega \models \neg \phi_i \\ 0 & \text{sinon} \end{cases} \quad (2.6)$$

En combinant toutes ces fonctions coûts k_i par l'addition, nous obtenons le coût d'une interprétation défini par :

$$k(\omega) = \sum_{i=1}^n k_i(\omega)$$

De manière similaire, la logique possibiliste peut être aussi considérée comme une logique multi-sources : chaque formule $\langle \phi_i, \beta_i \rangle$ d'une base de connaissances possibiliste provient d'une source indépendante et exprime une contrainte sur une distribution de possibilités π_i définie par :

$$\pi_i(\omega) = \begin{cases} 1 - \beta_i & \text{si } \omega \models \neg \phi_i \\ 1 & \text{sinon} \end{cases} \quad (2.7)$$

En combinant toutes ces distributions de possibilités π_i par le minimum, nous obtenons la distribution de possibilités la moins spécifique qui satisfait les contraintes :

$$\pi^*(\omega) = \min_{i=1 \dots n} \pi_i(\omega)$$

Si nous utilisons une autre norme triangulaire comme le produit, nous obtenons :

$$\pi^*(\omega) = \prod_{i=1 \dots n} \pi_i(\omega)$$

Le logarithme de cette distribution de possibilités correspond à l'opposé de la fonction coût induite par la base de connaissances à pénalités où les pondérations α_i sont telles que $\alpha_i = -\log(1 - \beta_i)$.

$$\log(\pi^*(\omega)) = \sum_{\omega \models \neg \phi_i} \log(1 - \beta_i) = -k(\omega) \quad (2.8)$$

2.5.5 La logique des pénalités et les fonctions de croyance

Dans [46], Dupin de Saint-Cyr a montré la relation entre la théorie des fonctions de croyances et la logique des pénalités sous forme de deux points :

- Le coût d'interprétation $k_{PK} : \Omega \rightarrow [0, +\infty]$ induit par une base à pénalités PK composée de n formules pondérées correspond à la fonction de contour Pl : $\Omega \rightarrow [0, 1]$ induite par la règle de combinaison de Dempster sur n fonctions supports (une pour chaque formule ϕ_i).

Soit $PK = \{(\phi_i, \alpha_i), i = 1, \dots, n\}$ une base de connaissances à pénalités. La fonction de masse m_i est définie comme suit :

$$m_i(\phi_i) = 1 - e^{-\alpha_i}$$

$$m_i(\top) = e^{-\alpha_i}$$

avec $e^{-\infty} = 0$ vu que $\alpha_i \in [0, 1]$.

Soit $m = m_1 \otimes \dots \otimes m_n$ le résultat de la combinaison des m_i avec la règle de Dempster [21] sans re-normalisation.

La fonction de contour pl : $\Omega \rightarrow [0, 1]$ associée à m est la restriction de la fonction de plausibilité Pl sur des singletons, c'est à dire que :

$$Pl(\omega) = Pl(\{\omega\}) = \sum_{\omega \models \phi} m(\phi)$$

Après développement, nous obtenons :

$$Pl(\omega) = e^{-K_{PK}(\omega)}$$

ou encore :

$$K_{PK}(\omega) = -\text{Log}(Pl(\omega))$$

Cette équation ne peut être étendue aux formules ; Il n'y a pas d'équivalence entre la fonction de plausibilité et la fonction coût K_{PK} .

- La fonction $K_{PK} : \mathcal{L} \rightarrow [0, 1]$ correspond à une mesure de plausibilité dans le cas d'une version infinitésimale de la théorie de Dempster-Shafer où les masses impliquées sont toutes infiniment proches de 0 ou de 1.

2.5.6 La logique des pénalités et les fonctions conditionnelles ordinales

Les fonctions conditionnelles ordinales [39] associent à chaque interprétation ω un nombre entier $k(\omega)$ qui représente le "degré d'anormalité" de ω . Cette fonction k s'étend aux formules de la même façon que dans la logique des pénalités [46], par :

$$k(\phi) = \min_{\omega \models \phi} k(\omega)$$

L'interprétation de $k(\phi)$ a trait aux probabilités infinitésimales :

- Si $k(\phi) = \alpha$ alors la probabilité de ϕ est de l'ordre de ϵ^α , où ϵ est infiniment petit.
- Si $\alpha = 0$ alors $\text{Pr}(\phi) \approx k$ où $k > 0$.

Spohn a défini également le degré d'anormalité de ω sachant ϕ par :

$$k(\omega | \phi) = k(\omega) - k(\phi)$$

puis :

$$k(\psi \mid \phi) = \min_{\omega \models \phi} k(\omega \mid \phi)$$

Si ϕ et ψ ne sont pas incompatibles alors $k(\phi \wedge \psi) = k(\phi) + k(\psi \mid \phi)$.

Le caractère additif de la fonction k n'a pas trait à une somme de coûts issue de la violation de formules (comme le cas de la logique des pénalités) mais à l'application du théorème de Bayes sur des probabilités infinitésimales.

Remarque 2.3 *Pour les fonctions conditionnelles ordinales, il n'existe pas de bases de connaissances sous-jacentes à la construction de la fonction k [38].*

2.6 Conclusion

Les différentes logiques présentées dans ce chapitre permettent de traiter les informations incertaines. Plusieurs auteurs se sont intéressés à faire le parallèle entre elles. La logique de probabilités, la logique des possibilités et les OCF admettent l'existence d'interprétations totalement possibles exprimées par la valeur 1 dans la théorie des possibilités et par 0 dans les OCF. Cette notion n'existe pas lorsque nous utilisons les relations de plausibilité (les fonctions de croyance). Les théories des probabilités et de possibilités admettent l'existence de mondes totalement impossibles exprimée par 0. Dans les autres formalismes, toutes les interprétations sont quelque peu possibles.

Dans ce chapitre, nous nous sommes intéressés, entre autres, à deux types de logique : la logique possibiliste quantitative et la logique des pénalités. Ces modes de représentation des connaissances incertaines ainsi que leurs modes d'inférence feront l'objet de notre étude dans les chapitres ultérieurs.

Nous nous intéressons dans le chapitre suivant aux modes de représentation graphiques et plus précisément

- aux réseaux causaux probabilistes,
- et aux réseaux causaux possibilistes qualitatifs

Les réseaux causaux possibilistes quantitatifs, faisant l'objet de notre étude, seront abordés plus en détails dans la partie consacrée à notre approche.

Chapitre 3

Modèles Graphiques de l'incertain : Représentation et Raisonnement

3.1 Introduction

Les approches logiques de l'incertain permettent d'utiliser un langage formel pour la description des connaissances et le raisonnement automatique. Elles constituent une référence aux autres formalismes surtout pour le raisonnement. Néanmoins, les connaissances ne sont pas structurées.

Plusieurs recherches ont justement permis l'émergence d'un certain nombre de modèles graphiques offrant un cadre de représentation plus structuré. En effet, afin de capturer les différents aspects inhérents au raisonnement incertain, il est primordial de tenir compte des influences pouvant exister entre certains événements et la répercussion de la prise en charge d'une nouvelle connaissance sur le réseau.

Parmi les modèles graphiques les plus répandus, nous distinguons : les réseaux causaux probabilistes [10, 11, 12], les arbres de décision [47], les diagrammes d'influence [48, 49, 50] et les *Valuation Based Systems (VBS)* [51, 52, 53, 54, 55]. L'ensemble de ces modèles graphiques reposent sur la théorie des probabilités.

Pour les mêmes raisons qu'en logique et dans le même ordre d'idée, des chercheurs ont été amenés à adapter les mécanismes de propagation probabilistes en proposant des versions possibilistes. Ainsi, dans [56], les auteurs proposent un algorithme de propagation possibiliste pour les hypergraphes. Dans [57], Fonck a adapté l'algorithme de Pearl [10] dans le contexte possibiliste. Dans [58, 59, 60] les auteurs ont développé une boîte à outils pour la propagation possibiliste pour les réseaux possibilistes indirects et plus récemment dans [61] l'auteur a adapté les algorithmes de propagation des réseaux causaux probabilistes aux réseaux causaux possibilistes basés sur le produit et sur le minimum.

Les VBS ont également été adaptés en théorie des possibilités. En effet, Shenoy [52] a proposé une version possibiliste de l'algorithme de propagation dans les VBS utilisant uniquement l'opérateur produit. Cano, Delgado et Moral [62] ont présenté un système de propagation pour des graphes simplement connectés utilisant des valuations. Une généralisation de cet algorithme au graphe avec boucles a été développé par Fonck [57]. Dans une représentation graphique, nous distinguons deux composantes :

- Une composante graphique matérialisée par des graphes pouvant être de différentes structures. Elle explicite l'ensemble des variables dont il est question ainsi que les liens qui existent entre elles, mettant ainsi en exergue les différentes in-

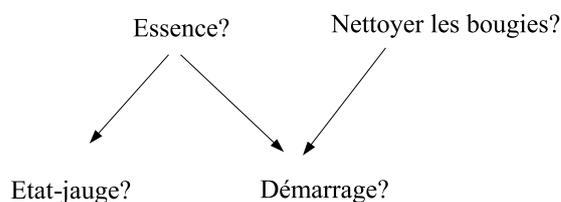


FIGURE 3.1 – Réseau causal correspondant à l'exemple 3.1

fluences entre les variables.

- Une composante numérique qui quantifie les différents liens qui existent entre les variables.

Les modèles possibilistes, basés sur la théorie des possibilités, offrent deux contextes de travail, permettant soit une modélisation qualitative soit une modélisation quantitative selon l'hypothèse de la problématique à traiter. Ceci permet en fait de définir deux types de réseaux causaux possibilistes : les réseaux causaux possibilistes basés sur le minimum (qualitatif) et les réseaux causaux possibilistes basés sur le produit (quantitatif).

Nous allons aborder dans ce chapitre les différentes notions inhérentes aux principaux modèles graphiques de l'incertain en commençant par les modèles numériques issus de la théorie des probabilités, puis les modèles issus de la théorie des possibilités.

3.2 Réseaux causaux Bayésiens

Les réseaux causaux bayésiens sont des modèles graphiques dirigés ou orientés. Ils sont issus d'un mariage entre la théorie des probabilités et la théorie des graphes apportant ainsi des outils naturels pour traiter le problème crucial de l'incertitude dans le domaine de l'intelligence artificielle. La théorie des probabilités est à la base de ces modèles graphiques assurant leurs consistences. La théorie des graphes fournit une interface intuitive permettant ainsi à un humain de structurer un problème par un ensemble de composantes interagissant entre elles.

Cette synthèse a pour but de donner un aperçu sur les principaux concepts des réseaux bayésiens et de présenter les principaux algorithmes exacts de la propagation.

3.2.1 Les concepts de base

Soit l'exemple suivant [63] extrait du *frame problem* [64] :

Exemple 3.1 *En début de journée, la voiture ne démarrait pas. Le starter étant en fonction, les causes de la panne peuvent être diverses : Vu que le starter fonctionne, cela induit que la batterie est chargée. La cause la plus probable est que l'essence a dû être dérobé durant la nuit ou encore que les bougies ne sont pas nettoyées. Cela peut être également dû au carburateur non nettoyé, ou encore à quelque chose de plus sérieux. Afin de localiser la panne, la jauge d'essence est d'abord vérifiée. Celle-ci indiquait la moitié du réservoir. Ainsi, la décision prise est de nettoyer les bougies.*

Afin d'implémenter ce type de raisonnement sur machine, il est nécessaire de pouvoir répondre à des questions de type : qu'est ce qui m'a permis d'affirmer que les causes : "Vol d'essence" et "bougies mal nettoyées" sont les plus probables ? Une autre question qui est posée est : qu'est ce qui m'a poussé à vérifier la jauge d'essence ?

Il est alors nécessaire de disposer d'outils afin de représenter le problème ainsi que des méthodes d'inférence.

Une façon de faire, consiste à construire un graphe représentant les relations causales entre les événements. Ainsi, cette situation peut être décrite par un graphe composé par des nœuds et des liens. Les événements sont représentés par les nœuds. Lorsque un nœud A a un impact causal sur le nœud B, alors les deux nœuds seront reliés par un lien.

La figure 3.1 illustre les relations causales de l'exemple 3.1.

3.2.1.1 Quelques définitions

Avant d'entamer les concepts de bases des réseaux causaux bayésiens, nous allons présenter, dans cette section, les préliminaires sur les graphes, nécessaires à la description des algorithmes d'inférence.

1. Généralités

Définition 3.1 (Graphe ou réseau) Soit V un ensemble fini de variables. Soit E une partie de $V \times V$. Un graphe noté $G=(V,E)$ est défini par deux ensembles V et E où V est un ensemble fini de nœuds $V=\{A, B, \dots\}$ et E est un ensemble de liens connectant des paires de nœuds dans V (voir figure 3.2b). Si les liens dans E sont orientés alors, ils sont dits des arcs et le graphe G associé est appelé graphe orienté.

Les figures (3.2a),(3.2b),(3.2c),(3.2d),(3.2e) représentent des graphes.

Définition 3.2 (Chaîne, Chemin) Soit $G=(V,E)$ un graphe. Une séquence de nœuds $[v_0, v_1, \dots, v_n]$ est une chaîne dans G si et seulement si :

$$\forall i, 1 \leq i \leq n, (v_{i-1}, v_i) \in E, (v_i, v_{i+1}) \in E$$

Soit $G=(V,E)$ un graphe orienté. Une séquence de nœuds $[v_0, v_0, \dots, v_n]$ est un chemin dans G si et seulement si :

$$\forall i, 1 \leq i \leq n, (v_{i-1}, v_i) \in E$$

Définition 3.3 (Cycle, Boucle) Soit $G=(V,E)$ un graphe. Un cycle est un chemin simple $[v_0, v_1, \dots, v_0]$.

Une boucle est un cycle indirect.

Définition 3.4 (Graphe Acyclique Orienté) Soit $G=(V,E)$ un graphe orienté. G est dit acyclique si G ne contient aucun cycle orienté.

Définition 3.5 (Parent, Enfant) Soit $G=(V,E)$ un graphe orienté. Si $(u,v) \in A$, alors u est un parent de v , et v est un enfant de u .

Définition 3.6 (Ascendant, Descendant) Soit $G=(V,E)$ un graphe orienté et u,v deux nœuds. S'il existe un chemin orienté de u à v alors u est un ascendant de v , et v est un descendant de u .

Définition 3.7 (Racine, Feuille) Une racine d'un graphe est un nœud sans parent. Une feuille d'un graphe est un nœud sans enfant.

Notations 3.1 – Parents : U_A est l'ensemble des parents du nœud A .
 – **Enfants :** Y_A est l'ensemble des enfants du nœud A .
 – **Descendants :** X_A est l'ensemble des descendants du nœud A .

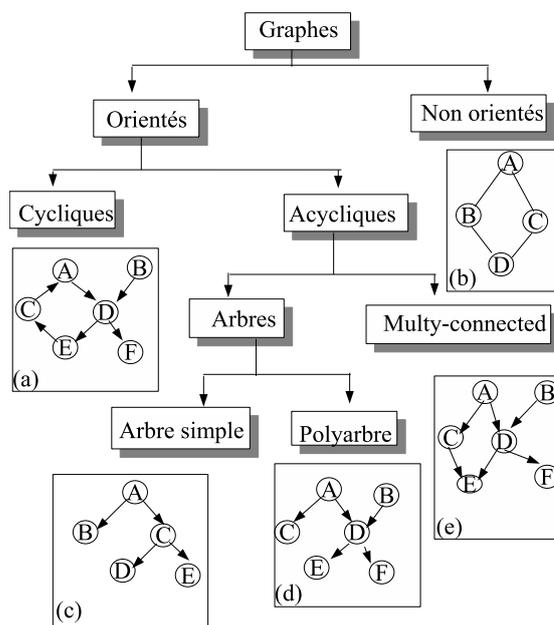


FIGURE 3.2 – Différents types de graphes

- **non-Descendants** : Z_A est l'ensemble des non-descendants du nœud A , tel que $Z_A = V - X_A$.

2. Les types de graphes

Selon la nature des liens qui existent entre les nœuds, plusieurs types de graphes ont été définis :

- **DAG** : (Directed Acyclic Graph) est un graphe orienté acyclique (sans cycle). Les figures (3.2c), (3.2d) et (3.2e) sont des DAG. La figure (3.2a) ne représente pas un DAG.
- **DAG simplement connecté** : c'est un DAG ne contenant pas de boucles. Un DAG est dit un arbre simple si chaque nœud a au plus un parent (voir figure (3.2c)). Autrement, il s'agit d'un polyarbre (voir figure (3.2d)).
- **DAG avec boucle** : est un DAG qui contient au moins une boucle. La figure (3.2e) contient la boucle ACED.

Remarque 3.1 les graphes orientés et non orientés sont très utilisés pour représenter des connaissances. Les modèles à base de graphes non dirigés appelés les modèles de Markov permettent de représenter des relations temporelles ou spatiales [65, 66, 67]. Les graphes dirigés acycliques sont utilisés afin de représenter des relations temporelles ou causales comme dans [68, 69, 70].

3.2.1.2 Les réseaux causaux et le critère de D-Séparation

Les réseaux causaux bayésiens sont issus des travaux de [10]. Ils sont fondés sur des Graphes Acycliques Dirigés (DAG) où les nœuds représentent des variables et les arcs résument les liens causaux qui existent entre elles. Les variables représentent des événements (propositions), pouvant avoir un nombre quelconque de valeurs.

Le raisonnement sous incertitude requiert la prise en charge de la retombée de la croyance d'un événement sur les croyances des autres événements du système. Un réseau causal, justement, permet d'étudier l'influence du changement sur le degré de certitude

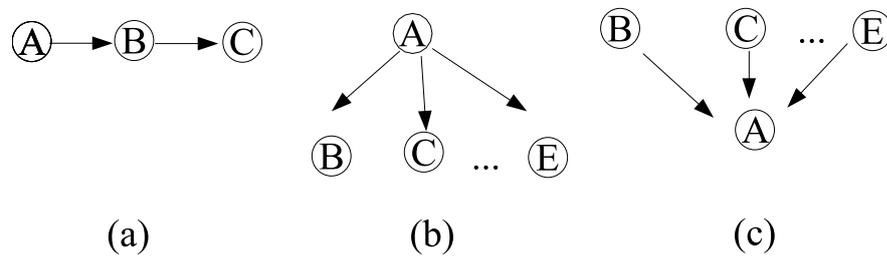
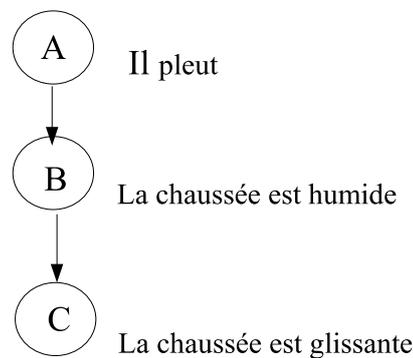


FIGURE 3.3 – Les types de liaisons dans un DAG



Une fois que B est connue, les variables A et C deviennent indépendantes

FIGURE 3.4 – Exemple de liaisons en série

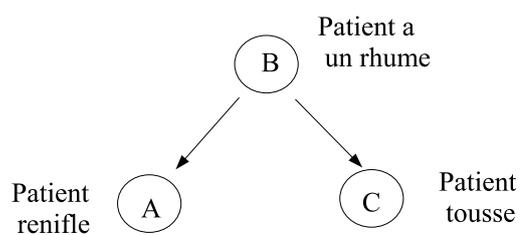
d'une variable sur l'ensemble des variable du réseau. Cette influence est en fonction de la topologie du DAG.

Les topologies des DAG : Il existe trois types de liaisons ou de connexions dans un DAG :

Connexions en série : Considérons la situation de la figure (3.3a). La variable A a une influence sur B, qui à son tour, influe sur C. D'une manière similaire, l'évidence sur C influe sur la certitude de A à travers B. D'autres part, si l'état de B est connu, alors le canal devient bloqué. A et C deviennent alors indépendants. A et C sont alors dits d-séparées étant donné B. Ainsi, l'évidence est transmise à travers la connexion en série jusqu'à ce que l'état de la variable en connexion soit connu.

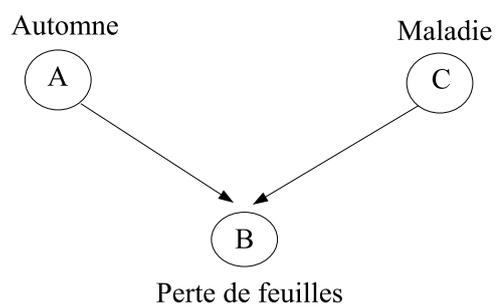
Connexions divergentes : La situation de la figure (3.3b) est une connexion divergente. L'influence passe entre l'ensemble des fils de A, jusqu'à ce que A soit connue. Ainsi, B,C,...,E sont d-séparées étant donné A. L'évidence peut être transmise à travers des connexions divergentes jusqu'à ce qu'elle soit instanciée.

Connexions convergentes : La figure (3.3c) illustre une connexion convergente. Si nous ne connaissons rien sur A excepté les connaissances inférées de ses parents B,...,E alors ces derniers sont séparés. Si A change alors la communication entre ses parents est alors ouverte.



A et C ne sont dépendantes que si B est inconnue

FIGURE 3.5 – Exemple de liaisons divergentes



A et C ne sont dépendantes que si B est connue

FIGURE 3.6 – Exemple de liaisons convergentes

Indépendance conditionnelle dans les DAGs : le critère de D-séparation La structure d'un DAG encode les relations d'indépendance et ceci sans tenir compte des valuations numériques. Pearl, Verma et Geiger [10, 71, 72, 73] ont examiné le problème de définir avec exactitude les types d'indépendance impliquées par une structure de DAG dans un réseau causal. Cette notion d'indépendance est à la base du critère de d-séparation. Le critère de d-séparation est une propriété purement graphique permettant de préciser dans quelles conditions une information peut être traitée localement sans perturber l'ensemble du graphe.

Ainsi, tout calcul d'incertitude dans une structure causale obéit au principe stipulant que si A et B sont d-séparées, alors une nouvelle information relative à l'une d'entre elles n'influe pas sur la certitude de l'autre. Cette caractéristique est très utile lors de la propagation d'une évidence e à travers un DAG.

La règle générale du critère de d-séparation est la suivante :

Définition 3.8 (Nœuds d-séparés par un ensemble de nœuds) Soit $G=(V,E)$ un DAG. Soient $Z \subseteq V$ et A et B deux nœuds $\in V-Z$. A et B sont dits d-séparés par Z si chaque chemin entre A et B est bloqué par Z . Plus précisément, si :

- La connexion est en série ou elle est divergente et l'état de Z est connu,
 - la connexion est convergente et ni Z ni aucun de ses descendants ne sont connus.
- Cette caractéristique est notée par :

$$\langle A \mid Z \mid B \rangle_G$$

Cette définition peut être étendue aux ensembles :

Définition 3.9 (ensembles) Soit $G=(V,E)$ un DAG. Soient X, Y, Z des sous-ensembles disjoints de V . X et Y sont d-séparés par Z si chaque nœud $X_i \in X$ et $Y_i \in Y$, X_i, Y_i sont d-séparés par Z .

Définition 3.10 (Chemin d-séparé par un ensemble de nœuds) Un chemin p est dit d-séparé par un ensemble Z de nœuds si et seulement si :

1. p contient une séquence en série $i \rightarrow m \rightarrow j$ ou une divergence $i \leftarrow m \rightarrow j$, telle que $m \in Z$, ou bien
2. p contient une convergence $i \rightarrow m \leftarrow j$ telle que $m \notin Z$ et telle qu'aucun descendant de m n'appartienne à Z .

Exemple 3.2 Dans le DAG représenté par la figure 3.7, les relations de d-séparation suivantes sont détectées :

- $X=\{X_2\}$ et $Y=\{X_3\}$ sont d-séparés par $Z=\{X_1\}$ car les deux chemins connectant X_2 à X_3 sont bloqués par Z
- Le chemin $X_2 \leftarrow X_1 \rightarrow X_3$ est bloqué car il s'agit d'une divergence dans laquelle le nœud milieu X_1 appartient à Z .
- Le chemin $X_2 \rightarrow X_4 \leftarrow X_3$ est bloqué car il s'agit d'une convergence dans laquelle le nœud X_4 ainsi que tous ses descendants n'appartiennent pas à Z .

Par contre $Z'=\{X_1, X_5\}$ ne d-sépare pas X_2 et X_3 car le chemin $X_2 \rightarrow X_4 \leftarrow X_3$ n'est pas bloqué par Z' vu que X_5 , qui est un descendant du nœud milieu X_4 , appartient à Z' . Ainsi, le fait de connaître l'effet X_5 rend ses causes X_2 et X_3 dépendantes.

Remarque 3.2 Le concept de d-séparation est utilisé afin de faciliter la mise à jour des probabilités dans un réseau causal après l'arrivée d'une nouvelle connaissance. Cette opération est aussi connue sous les noms d'inférence ou de propagation. Néanmoins, comme nous le verrons dans les sections suivantes, même en tenant compte de ce critère, le calcul d'inférence probabiliste est en général NP-Complexe.

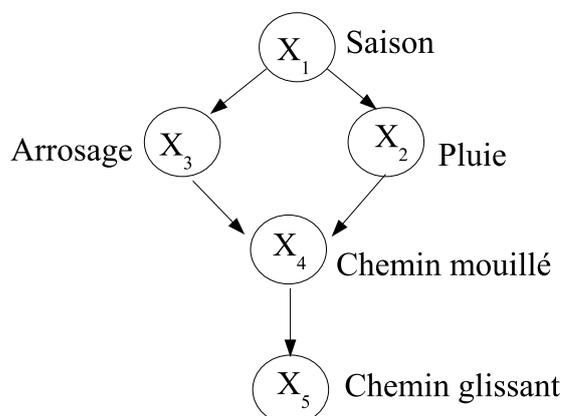


FIGURE 3.7 – Réseau bayésien représentant les dépendances entre cinq variables

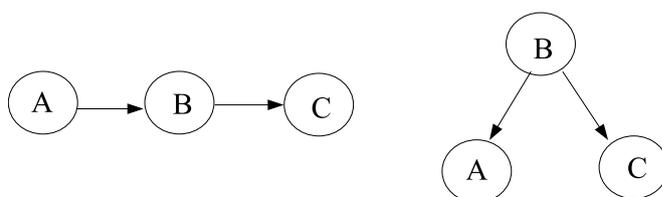


FIGURE 3.8 – Topologies où les nœuds A et C sont conditionnellement indépendants étant donné B

3.2.1.3 L'indépendance conditionnelle dans les DAGs

Le concept d'indépendance conditionnelle reflète la notion de blocage d'influence entre variables.

Définition 3.11 *X et Y sont indépendants conditionnellement de Z, noté $X \perp Y \mid Z$, si :*

$$P(X \mid Z, Y) = P(X \mid Z)$$

Remarque 3.3 *La notion d'indépendance conditionnelle apparaît dans les cas de connexions en série et de connexions divergentes (voir figure 3.8).*

Le lien entre le critère de d-séparation et d'indépendance conditionnelle est établi grâce au théorème de Verma et Pearl [74, 75] :

Théorème 3.1 *Soit $G=(V,E)$ un DAG. Soient $X \subset V$, $Y \subset V$ et $Z \subset V$, trois sous ensembles de nœuds. Si X et Y sont d-séparés dans G par Z, alors X et Y sont indépendants conditionnellement à Z, c.a.d, $X \perp Y \mid Z$.*

3.2.1.4 Modélisation d'un réseau Bayésien

Un réseau Bayésien permet de représenter un ensemble de variables aléatoires pour lesquelles nous connaissons un certain nombre de relations de dépendances et où une distribution de probabilités définie sur l'ensemble des variables est disponible. Formellement :

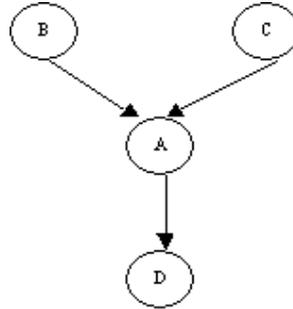


FIGURE 3.9 – Exemple d'un réseau Bayésien

Définition 3.12 Soit $G=(V,E)$ un DAG. Soit P la distribution de probabilités jointe sur l'ensemble des éléments de V . (G,P) est dit un réseau Bayésien si chaque variable $A \in V$ est conditionnellement indépendante de ses non-descendants Z_A étant donné ses parents (U_A) .

Il est à noter qu'en pratique, la distribution de probabilités jointe sur l'ensemble des variables n'est quasiment jamais spécifiée. Ainsi, ce sont en général des distributions de probabilités locales qui sont utilisées.

Spécification d'un réseau Bayésien : La construction d'un modèle Bayésien se fait en deux phases :

1. **Spécification de la composante qualitative :** Cette phase consiste à définir les relations d'influence pouvant exister entre les variables prises deux à deux. Ces influences seront représentées graphiquement par des arcs.
2. **Spécification de la composante quantitative :** Cette phase consiste à quantifier numériquement les différents liens qui existent dans le DAG par des probabilités conditionnelles de chaque variable A dans le contexte de ses parents (U_A) comme suit :
 - Si $(U_A) = \emptyset$ (A est un nœud racine) alors, il s'agit de spécifier les probabilités a priori relatives aux différentes instances de la variable A , tout en respectant la condition de normalisation qui stipule que :

$$\sum_a P(a) = 1$$

- Si $(U_A) \neq \emptyset$ alors il faudrait spécifier les probabilités conditionnelles des différentes instances a de A dans le contexte des différentes instances de ses parents u_A telles que :

$$\sum_a P(a | u_A) = 1$$

Ces spécifications numériques sont souvent données par des experts du domaine modélisé.

Exemple 3.3 Soit le DAG illustré dans la figure (3.9). Les distributions a priori sont données par la table (3.1) et les distributions conditionnelles sont données par les tables (3.2) et (3.3).

TABLE 3.1 – Les probabilités a priori P(B) et P(C)

B	P(B)	C	P(C)
b	.7	c	.5
$\neg b$.3	$\neg c$.5

TABLE 3.2 – Les probabilités conditionnelles P(A | B ∧ C)

A	B	C	P(A B ∧ C)	A	B	C	P(A B ∧ C)
a	b	c	.8	$\neg a$	b	c	.2
a	b	$\neg c$.6	$\neg a$	b	$\neg c$.4
a	$\neg b$	c	.9	$\neg a$	$\neg b$	c	.1
a	$\neg b$	$\neg c$.3	$\neg a$	$\neg b$	$\neg c$.7

La règle de chaînage : La règle de chaînage permet une représentation compacte de la distribution de probabilité associée à l'ensemble des variables. En effet, si nous considérons l'ensemble V, la représentation en mémoire de P(V) serait très fastidieuse. Par exemple si l'ensemble V est composé de 30 variables prenant chacune trois valeurs, alors il faudrait mémoriser plus de 10^{14} valeurs.

Afin de palier à cet inconvénient, la règle de chaînage permet, justement, de calculer P(V) à partir des probabilités conditionnelles spécifiées dans le réseau :

Définition 3.13 (Règle de chaînage) *La distribution de probabilités globale jointe à travers un ensemble de variables $V=\{A_1, \dots, A_N\}$ est exprimée par le produit des probabilités conditionnelles via la règle de chaînage suivante :*

$$P(A_1, \dots, A_N) = \prod_{i=1..N} P(A_i | U_{A_i}) \quad (3.1)$$

où U_{A_i} représentent les parents de A_i .

Exemple 3.4 *Soit le réseau causal de la figure 3.9 auquel est associé la table des probabilités a priori (table 3.1) et les tables des probabilités conditionnelles (tables 3.2 et 3.3).*

La distribution des probabilités jointe est définie par :

$$\forall a_i, b_i, c_i, d_i, P(a_i \wedge b_i \wedge c_i \wedge d_i) = P(a_i | b_i \wedge c_i) * P(d_i \wedge a_i) * P(b_i) * P(c_i)$$

où a_i, b_i, c_i, d_i représentent respectivement les instances des variables A, B, C, D.

Par exemple $P(\neg a \wedge b \wedge \neg c \wedge d) = P(\neg a | b \wedge \neg c) * P(d | \neg a) * P(b) * P(\neg c) = .4 * .1 * .7 * .5 = .014$.

3.2.2 Algorithmes de propagation dans les réseaux causaux bayésiens

3.2.2.1 Introduction

Pour les réseaux causaux probabilistes, l'inférence consiste à propager une évidence (une nouvelle information) à travers un réseau et de voir son impact. La propagation entraîne en fait, une mise à jour des distributions de probabilités associées aux variables après l'arrivée de l'évidence.

Il existe principalement deux types d'algorithmes d'inférence :

1. **Algorithmes exacts :** Ces algorithmes calculent les distributions de probabilités associées aux nœuds exactement. L'algorithme exact le plus fondamental est celui proposé par Kim et Pearl [76] et Pearl [10, 77]. L'impact de chaque évidence

TABLE 3.3 – Les probabilités conditionnelles $P(D | A)$

D	A	$P(D A)$
d	a	.6
d	$\neg a$.1
$\neg d$	a	.4
$\neg d$	$\neg a$.9

est perçu comme une perturbation se propageant en parallèle à travers le réseau, via un mécanisme de passage des messages entre les variables voisines. Dans le cas d'un graphe causal simplement connecté, l'algorithme converge rapidement. Néanmoins, pour le cas d'un réseau avec boucles, l'algorithme ne donnera pas les résultats escomptés vu la présence de boucles. Pour cela d'autres algorithmes, tel que celui de l'arbre de jonction [11], ont été développés nécessitant la transformation du graphe initial en un arbre de jonction formé d'un ensemble de groupes de variables.

2. **Algorithmes approximatifs** : Ce type d'algorithme est utilisé dans le cas où il n'existe pas d'algorithmes exacts ou encore lorsque ceux-ci présentent une complexité calculatoire intense. En effet, les algorithmes exacts sont NP-Complet [78] excepté pour les polyarbres.

L'idée de base de ces méthodes est de générer un échantillon à partir de la distribution de probabilités jointe des variables, puis d'utiliser l'échantillon généré pour calculer les valeurs de probabilités d'événement après prise en compte de l'évidence [79].

Les méthodes de propagation approximatives peuvent être classées en deux types : les méthodes de simulation stochastiques et les méthodes de recherche déterministes. Les méthodes du premier type génèrent l'échantillon à partir de la distribution de probabilités jointe en utilisant des mécanismes aléatoires. Pour le deuxième type de méthodes, l'échantillon est généré d'une manière systématique [79]. Parmi les algorithmes approximatifs, nous distinguons *loopy belief propagation* [80].

Dans ce qui suit, nous présentons les algorithmes exacts les plus utilisés selon la structure du DAG (polyarbre, réseau causal avec boucles).

3.2.2.2 La propagation dans les Polyarbres

Nous allons présenter dans cette section une version centralisée de l'algorithme de Pearl [76, 10] pour la propagation probabiliste dans un polyarbre proposé dans [81]. Cet algorithme a la particularité d'être de complexité polynômiale.

Le principe de propagation : Dans un polyarbre, deux nœuds sont reliés par un chemin unique. Ceci induit que chaque nœud X_i divise le polyarbre en deux parties : l'une incluant les nœuds accessibles à partir des parents du nœud X_i et l'autre inclut les nœuds accessibles à partir des enfants. Afin de faciliter le calcul, l'évidence E est décomposée en deux sous ensembles disjoints E_A^- et E_A^+ (voir la figure (3.10)) correspondants aux deux sous graphes séparés par le nœud A dans le polyarbre. Ces deux ensembles sont définis par :

- $E_A^- = \{E_A, E_{AY_1}^-, \dots, E_{AY_m}^-\}$ où E_A correspond au nœud A et $E_{AY_j}^-$ correspond à l'instanciation des variables en-dessous de l'arc de A à Y_j .
- $E_A^+ = \{E_{U_1A}^+, \dots, E_{U_nA}^+\}$ où $E_{U_iA}^+$ correspond à l'instanciation des variables au-dessus de l'arc de U_i à A .

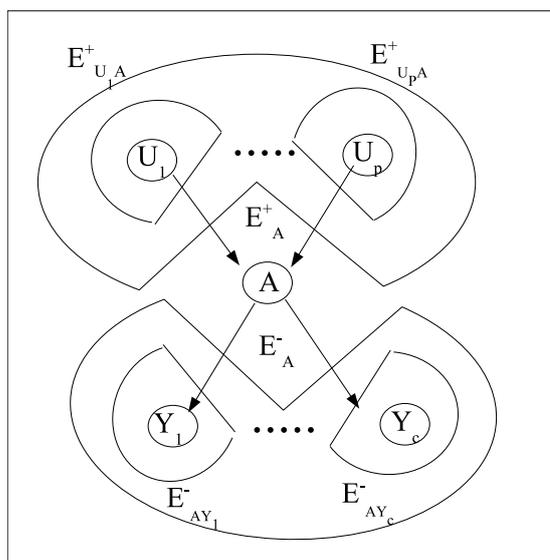


FIGURE 3.10 – Cautionnement de l'évidence E en sous-ensembles associés aux parents et aux enfants d'un nœud typique A

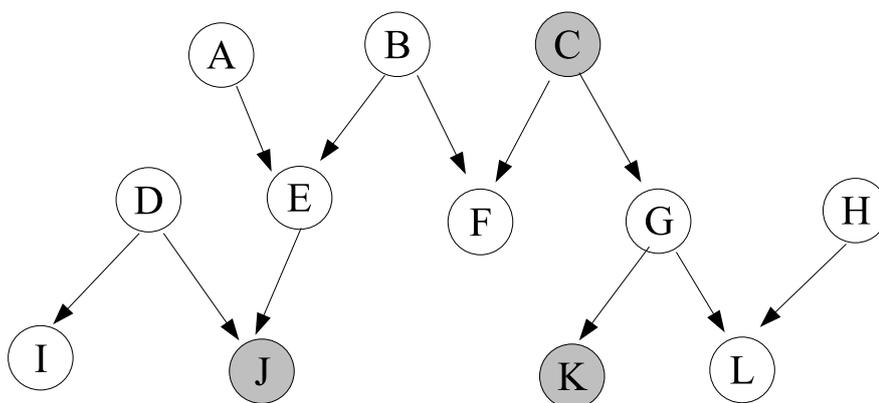


FIGURE 3.11 – Exemple d'un polyarbre

e_A^+ et e_A^- représentent respectivement l'évidence attachée au nœud dans E_A^+ et E_A^- . e_A dénote l'évidence attachée au nœud A .

Exemple 3.5 *Considérons le DAG représenté par la figure (3.11). Supposons que les variables binaires suivantes sont instanciées : $A=\neg a, B=b, D=\neg d, G=1, K=\neg k$. Alors, $E = \{A, B, D, G, K\}$, $E_A^+ = \{E_{BA}^-, E_{CA}^-, E_{EA}^-\} = \{B, D\}$, $E_A^- = \{E_A, E_{AG}^-, E_{AH}^-\} = \{A, G, K\}$, $e = \{\neg a, b, \neg d, g, \neg k\}$, $e_A^+ = \{e_{BA}^-, e_{CA}^-, e_{EA}^-\} = \{b, \neg d\}$, $e_A^- = \{e_A, e_{AG}^-, e_{AH}^-\} = \{\neg a, g, \neg k\}$*

La propagation de l'évidence est effectuée d'une manière efficace en combinant les informations provenant de différents sous-graphes en utilisant la technique de passage de message (calculs locaux) d'un sous graphe vers un autre. Plus précisément, les communications entre les variables se font en utilisant deux types de messages :

- λ message pour transmettre l'information des enfants vers les parents,
- μ message¹ pour transmettre l'information des parents vers les enfants.

1. Ce message est noté par μ au lieu de π (comme dans la littérature), afin de ne pas confondre

D'une manière formelle [10], chaque nœud A sera ainsi caractérisé par sa probabilité conditionnelle en se basant sur une évidence totale e , notée $Bel(A)$, qui sera fonction des valeurs λ et μ relatives à ce nœud.

La valeur $\lambda(A)$ est calculée à partir des λ messages reçus à partir des enfants du nœud A et la valeurs $\mu(A)$ est calculée à partir des μ messages reçus à partir des parents du nœud A.

Ainsi, $\forall a \in D_A$:

- La mesure de probabilité courante de chaque instance a basée sur l'évidence totale e est définie par :

$$Bel(a) = P(a | e) = \alpha * \lambda(a) * \mu(a) \quad (3.2)$$

où $\alpha = \frac{1}{\sum_a Bel(a)}$ est le facteur de normalisation

- La valeur λ associée à chaque instance $a \in D_A$ est définie par :

$$\lambda(a) = P(a | e_A^-) = \lambda_A(a) * \prod_{j=1}^m \lambda_{Y_j}(a) \quad (3.3)$$

où $\lambda_A(a)$ dénote l'évidence totale relative au nœud A telle que :

$$\lambda_A(a) = \begin{cases} 0 & \text{si } e_A \neq a \text{ (A est instanciée à } a \text{ avec } e_A \neq a) \\ 1 & \text{sinon (A est instanciée à } a \text{ avec } e_A = a \text{ ou A n'est pas instanciée)} \end{cases} \quad (3.4)$$

- La valeur μ associée à chaque instance $a \in D_A$ est définie par :

$$\mu(a) = P(a | e_A^+) = \sum_u P(a | u) * \prod_{i=1}^n \mu_A(u_i) \quad (3.5)$$

- Le message λ de A vers ses parents U_i , ($i \in \{1, \dots, n\}$) où $U_i = u_i$ est défini par :

$$\lambda_A(u_i) = P(e_{U_i A}^- | u_i) = \beta * \sum_a \lambda(a) \left[\sum_{u_k: k \neq i} P(a | u) * \prod_{k \neq i} \mu_A(u_k) \right] \quad (3.6)$$

où β est une constante de normalisation.

Remarque 3.4 Dans le cas où chaque nœud dans le graphe a un parent unique alors la valeur λ est simplifiée comme suit :

$$\lambda_A(u_i) = \beta * \sum_a \lambda(a) * P(a | u)$$

- Le message μ de A à ses enfants Y_j , ($j \in \{1, \dots, m\}$) où $A=a$ est défini par :

$$\mu_{Y_j}(a) = P(a | e_{AY_j}^+) = \alpha * \lambda_A(a) * \prod_{i=1, \dots, m, i \neq j} \lambda_{Y_i}(a) * \mu(a) \quad (3.7)$$

Algorithme de propagation L'algorithme de propagation de base pour les polyarbres a été proposé par Pearl dans [10]. Néanmoins, Peot et Shachter [81] ont proposé une version révisée de l'algorithme sans modifier sa complexité. Cette version converge au bout de deux itérations.

Algorithme 3.1 : Propagation dans un polyarbre

Début

avec la distribution de possibilités.

Choix de la racine de propagation Cette phase consiste à choisir un nœud arbitraire parmi l'ensemble le plus restreint des nœuds connectés contenant les nœuds observés noté \mathcal{S} .

Initialisation Dans cette étape, tous les λ et μ valeurs et messages sont initialisés comme suit :

1. Pour chaque nœud A racine, la valeur $\mu(a)$ est initialisée à $P(a)$, $\forall a \in D_A$.
2. Pour chaque nœud A observé,
 - Si A est instanciée à a , alors $\lambda(a)$ est initialisée à 1,
 - sinon $\lambda(a)$ est initialisée au degré 0.

Collecte de l'évidence Lors de cette phase, pour chaque nœud dans \mathcal{S} :

- ses valeurs λ et μ sont calculées en utilisant respectivement les équations (3.3) et (3.5) permettant ainsi le passage de message à son nœud adjacent dans la direction du pivot.

Si ce nœud est un parent alors

- le message est calculé en utilisant l'équation (3.6) **Sinon**

Le message est calculé en utilisant l'équation (3.7)

- la collecte de l'évidence débute avec le nœud le plus lointain du pivot dans \mathcal{S} .

Distribution de l'évidence Lors de cette phase, chaque nœud :

- calcule ses valeurs λ et μ en utilisant respectivement les équations (3.3) et (3.5)
- Passe les messages aux nœuds adjacents en commençant du pivot lui même jusqu'à atteindre les feuilles.

Le calcul des messages à envoyer aux nœuds adjacents dépend de son type :

Si le nœud traité est un parent alors

- le message est calculé en utilisant l'équation (3.6)

Sinon

Le message est calculé en utilisant l'équation (3.7)

Marginalisation Pour chaque nœud A , $Bel(A) = P(a | e)$ est calculé en utilisant l'équation(3.2).

Fin

Exemple 3.6 Considérons le DAG donné par la figure (3.11). Supposons que de nouvelles informations ont été obtenues pour les nœuds C, J et K . Ainsi, l'ensemble $\mathcal{S} = \{J, E, B, F, C, G, K\}$. Supposons que le nœud pivot est F . La figure (3.12a) représente le diagramme correspondant. Les différents passages de messages sont résumés par la figure (3.12b).

3.2.2.3 La propagation dans les graphes à connexions multiples

L'algorithme de propagation pour les polyarbres décrit par Pearl [10] n'est pas adapté pour les graphes à connexions multiples. Dans [12], les auteurs proposent un algorithme d'inférence exacte plus efficace et générale (pouvant être appliqué aussi bien pour les polyarbres que pour les graphes à connexions multiples), dit algorithme de l'arbre de jonction, qui consiste à compiler le graphe original en une structure d'arbre secondaire appelée arbre de jonction, puis de réaliser la transmission de message sur cet arbre. Cet algorithme fût amélioré par la suite par Jensen [11]. Ainsi, les experts formalisent leurs connaissances en utilisant les réseaux causaux et la propagation s'effectue sur une structure secondaire (arbre de jonction) caractérisée par :

- Un arbre de jonction (Junction Tree) \mathcal{JT} est formé d'un ensemble de nœuds. Chaque nœud est composé par un ensemble de variables ou *clusters* notés C_i .

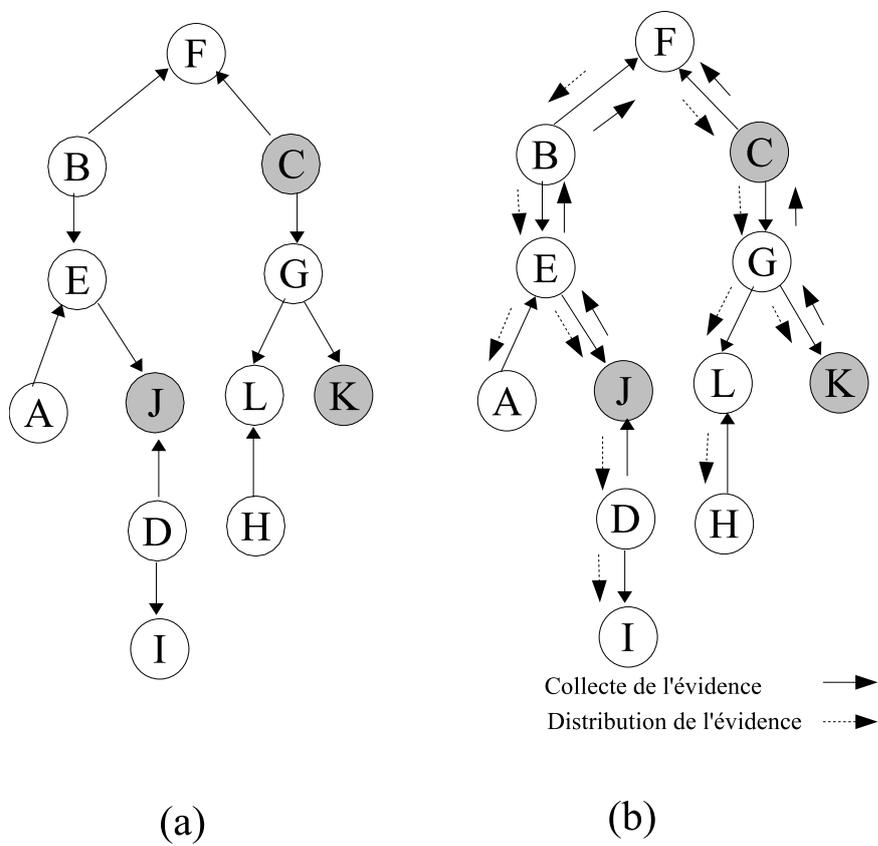


FIGURE 3.12 – Exemple de propagation dans un DAG simplement connecté

- Chaque arc est étiqueté par l'intersection des deux clusters adjacents C_i et C_j appelé séparateur et noté par S_{ij} .
- A chaque cluster C_i (respectivement séparateur S_{ij}) de \mathcal{JT} est affecté une distribution jointe locale relative aux variables du cluster (respectivement au séparateur), appelée potentiel, noté par ψ_{C_i} (respectivement $\psi_{S_{ij}}$).
- Une distribution de probabilités jointe globale unique est associée à un arbre de jonction \mathcal{JT} . Elle est définie par :

Définition 3.14 *La distribution jointe associée à un arbre de jonction \mathcal{JT} est :*

$$P_{\mathcal{JT}}(A_1, \dots, A_N) = \frac{\prod_{i=1}^m \psi_{C_i}}{\prod_{j=1}^{m-1} \psi_{S_{ij}}} \quad (3.8)$$

où m représente le nombre de clusters dans l'arbre de jonction.

- **Définition 3.15 (Lien consistant)** *Soient C_i et C_j deux clusters adjacents dans un arbre de jonction \mathcal{JT} et S_{ij} le séparateur associé. Le lien entre C_i et C_j est dit consistant ou stable si :*

$$\sum_{C_i \setminus S_{ij}} \psi_{C_i} = \psi_{S_{ij}} = \sum_{C_j \setminus S_{ij}} \psi_{C_j} \quad (3.9)$$

Si tous les liens dans l'arbre de jonction sont consistants, alors l'arbre de jonction est dit globalement consistant.

Si un arbre de jonction est globalement consistant, alors le potentiel associé à chaque cluster C_i doit satisfaire :

$$\psi_{C_i} = P(C_i) \quad (3.10)$$

- La distribution de probabilité associée à chaque variable A contenue dans un cluster C_i est obtenue en marginalisant ψ_{C_i} sur A comme suit :

$$P(A) = \sum_{C_i \setminus A} \psi_{C_i} \quad (3.11)$$

Principalement, l'algorithme d'inférence exact pour les graphes à connexions multiples se comporte de la manière suivante :

- **La phase de construction :** Elle nécessite un ensemble de sous-étapes afin de transformer le graphe initial en un arbre de jonction, dont les nœuds sont des **clusters** (regroupements des nœuds du graphe initial). Cette transformation est nécessaire pour :

- Éliminer les boucles du graphe initial,
- Optimiser l'algorithme d'inférence.

Elle s'effectue en trois étapes :

1. Moralisation du DAG initial \mathcal{G} ,
2. Triangulation du graphe moral,
3. Création de l'arbre de jonction à partir du graphe moral triangulé.

- **La phase de propagation :** C'est la phase de calcul de l'ensemble des distributions de probabilités du réseau après l'avènement de nouvelles informations. Cette phase s'effectue par le biais de passage de messages entre les nœuds de l'arbre de jonction.

La phase de construction comporte les étapes suivantes :

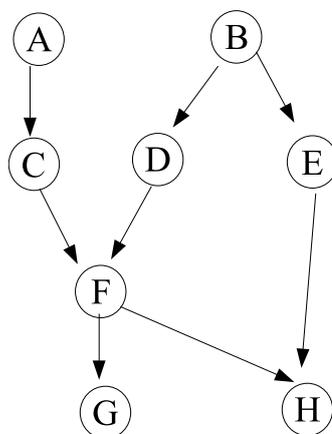


FIGURE 3.13 – Un graphe acyclique dirigé (DAG)

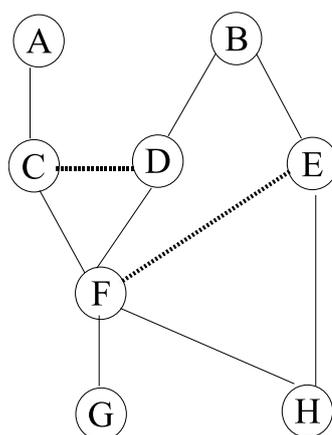


FIGURE 3.14 – Graphe moral associé au DAG de la figure 3.13

Les transformations graphiques : Cette étape consiste donc à construire l'arbre de jonction associé au réseau causal original \mathcal{G} .

Étape 1 : Moralisation : Elle consiste à marier deux à deux les parents de chaque variable en les reliant par un arc non-dirigé. A partir d'un graphe \mathcal{G} , le graphe moral associé noté $\mathcal{G}_{\mathcal{M}}$ est obtenu par la procédure suivante :

Algorithme 3.2 : Construction du graphe moral [12, 82]

Début

Associer au DAG initial \mathcal{G} un graphe non dirigé en éliminant les directions de tous les arcs,

Construire le graphe moral $\mathcal{G}_{\mathcal{M}}$ à partir de \mathcal{G} en reliant les parents de chaque nœud en rajoutant des arcs non dirigés.

Fin

Exemple 3.7 La figure 3.14 représente le graphe moral associé au DAG de la figure 3.13. Les arcs en pointillés représentent les arcs qui ont été ajoutés.

Étape 2 : Triangulation du graphe moral :

En général, il existe plusieurs façons de trianguler un graphe moral. La tâche de trouver la triangulation la plus optimale est NP-complet [83, 84, 78]. Néanmoins, plusieurs heuristiques ont été développées afin de réduire le coût induit par l'inférence, parmi lesquelles se trouve le critère de sélection de nœuds qui est une heuristique (induisant une complexité polynômial) produisant une triangulation de haute qualité [85].

La procédure de triangulation qui suit permet également d'identifier l'ensemble des clusters ou cliques (noté *cluster-set*) en utilisant l'algorithme de Golumbic [86] qui garantit qu'aucun cluster n'est inclus dans un autre cluster plus grand.

Algorithme 3.3 : Triangulation du graphe moral et identification des clusters [85]

ters [85]

Debut

Cluster-set := \emptyset ;

$\mathcal{G}'_{\mathcal{M}}$:= $\mathcal{G}_{\mathcal{M}}$;

Tantque il existe des nœuds dans $\mathcal{G}'_{\mathcal{M}}$ **Faire**

Si il existe un nœud A tel que tous ses nœuds adjacents sont connectés

Alors

Créer un cluster C contenant A ainsi que ses nœuds adjacents ;

Supprimer A de $\mathcal{G}'_{\mathcal{M}}$

Sinon

Sélectionner un nœud A ayant le nombre de nœuds adjacents le plus petit ;

Ajouter des arcs afin de connecter l'ensemble des nœuds adjacents ;

Pour chaque arc ajouté à $\mathcal{G}'_{\mathcal{M}}$, ajouter l'arc correspondant à $\mathcal{G}_{\mathcal{M}}$;

Créer un cluster C contenant A ainsi que ses nœuds adjacents ;

Supprimer A de $\mathcal{G}'_{\mathcal{M}}$

Si $C \notin$ *cluster-set* **Alors** *cluster-set* := *cluster-set* \cup $\{C\}$;

Fin

Exemple 3.8 La figure (3.15) représente le graphe triangulé obtenu à partir du graphe moral de la figure (3.14). L'arc en pointillé représente l'arc ajouté par la procédure de triangulation.

L'ordre d'élimination des variables ainsi que la construction des cliques sont illustrés par la table (3.4).

Étape 3 : Construction de l'arbre de jonction optimal : Afin de construire l'arbre de jonction, il faudrait relier les différents clusters générés par la procédure précédente. Cette tâche est réalisée en insérant, itérativement, des séparateurs entre des paires de cliques, jusqu'à ce que les cliques soient connectées par $m-1$ séparateurs, où m représente le nombre de clusters [87].

La procédure correspondante est donnée en deux parties : la première partie consiste à présenter la construction de l'arbre de jonction et la seconde étape est relative aux choix appropriés des séparateurs.

TABLE 3.4 – L'ordre d'élimination des variables

Nœuds éliminés	Clusters induits	Arcs ajoutés	Cluster-set
A	AC	rien	{AC}
G	GH	rien	{AC, GF}
C	CDF	rien	{AC, GF, CDF}
H	EFH	rien	{AC, GF, CDF, EFH}
B	BDE	(D,E)	{AC, GF, CDF, EFH, BDE}
D	DEF	rien	{AC, GF, CDF, EFH, BDE, DEF}
E	EF	rien	{AC, GF, CDF, EFH, BDE, DEF} (EF \subseteq DEF)
F	F	rien	{AC, GF, CDF, EFH, BDE, DEF} (F \subseteq DEF)

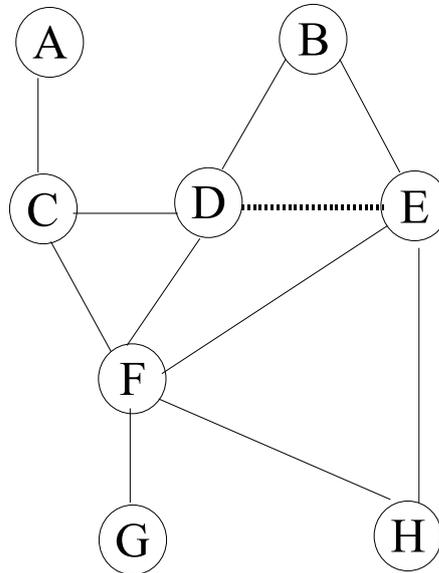


FIGURE 3.15 – Triangulation du graphe moral de la figure 3.14

1. **Construction de l'arbre de jonction optimal :** La procédure suivante permet de construire un arbre de jonction optimal à partir d'un ensemble de m clusters (cluster-set) et ceci en les reliant par un ensemble de séparateurs (Separator-set) :
Algorithme 3.4 : Construction d'un arbre de jonction optimal[88]

Debut

Pour $i := 1$ à m **Faire**

Separator-set := \emptyset ;

$C_i := \text{Cluster-set}[i]$;

Pour $j := 1$ à $(m-1)$ **Faire**

$C_j := \text{Cluster-set}[j]$;

Créer un séparateur candidat S_{ij} pour les deux clusters C_i et C_j ;

Insérer S_{ij} dans *Separator-set* ;

Sélectionner un séparateur S_{ij} à partir de *Separator-set*, en accord avec le critère de sélection spécifié ci-après ;

Insérer le séparateur S_{ij} entre le cluster C_i et le cluster C_j

Fin

2. **Choix des séparateurs appropriés :** Dans le but de décrire comment s'effectue la sélection du séparateur, les notions suivantes sont définies :
- Le poids associé à un nœud A est la taille de son domaine représentée par le nombre de ses valeurs.
 - Le poids associé à un cluster est le produit des poids des variables qui le constituent.
 - La masse d'un séparateur S_{ij} est le nombre de variables qu'il contient ou c'est le nombre de variables contenues dans $C_i \cap C_j$.
 - Le coût d'un séparateur S_{ij} est le produit des poids de ses variables.
- Ainsi, le choix du séparateur dans l'algorithme précédent se fait comme suit :
- Le séparateur candidat doit avoir la masse la plus élevée.
 - Dans le cas où plusieurs séparateurs ont la même masse, le temps de l'inférence peut être optimisé en choisissant le séparateur induisant le moindre coût.

Exemple 3.9 *Considérons l'ensemble de clusters suivant : cluster-set = {AC, GF, CDF, EFH, BDE, DEF} obtenu à partir de la figure (3.15) de l'exemple précédent et supposons que les variables sont binaires. L'application du critère de sélection est résumée dans la table 3.5. Ainsi, l'ensemble des séparateurs associés aux différentes cliques est : separator-set = {C, DF, DE, EF, F}. L'arbre de jonction associé est représenté par la figure (3.16).*

TABLE 3.5 – Choix des séparateurs appropriés

Clusters	Séparateurs candidats	Masse	Coût sélectionnés	Séparateurs
AC,CDF	C	1	2	*
AC,BDE	\emptyset	-	-	
AC,DEF	\emptyset	-	-	
AC,EFH	\emptyset	-	-	
AC,GF	\emptyset	-	-	
CDF,BDE	D	1	2	
CDF,DEF	DF	2	4	*
CDF,EFH	F	1	2	
CDF,GF	F	1	2	
BDE,DEF	DE	2	4	*
BDE,EFH	E	1	2	
BDE,GF	\emptyset	-	-	
DEF,EFH	EF	2	4	*
DED,GF	F	1	2	
GF,EFG	F	1	2	*

La phase de propagation : Après avoir construit l'arbre de jonction optimal associé au graphe causal original, l'étape suivante consiste à calculer la composante numérique associée à l'arbre de jonction. Le processus général de propagation dans les graphes à connexions multiples sans variables d'observation est illustré par la figure (3.17). Il consiste à calculer la distribution de probabilités $P(A)$ associée à une variable A dans le contexte d'aucune évidence [88].

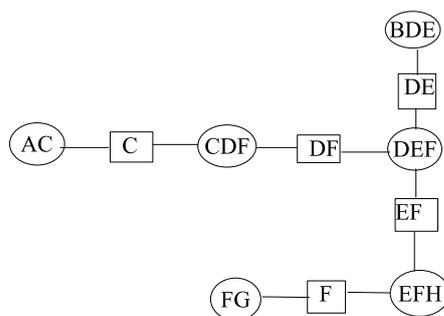


FIGURE 3.16 – Arbre de jonction associé au DAG de la figure 3.13

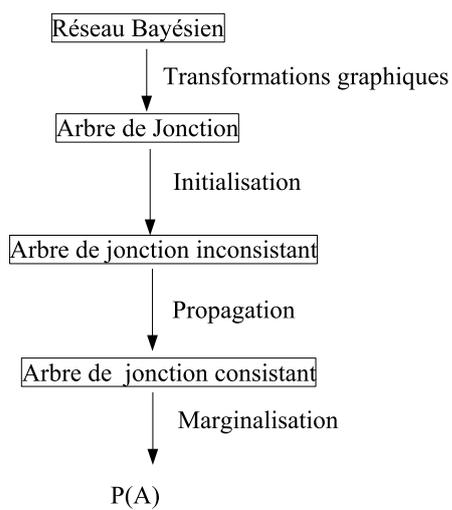


FIGURE 3.17 – Le diagramme de l'algorithme de propagation sans évidence

Notations 3.2 Le potentiel associé à un cluster C_i lors de la phase d'initialisation (respectivement lors de consistance globale) est noté $\psi_{C_i}^I$ (respectivement $\psi_{C_i}^C$.)

D'une manière détaillée, la propagation s'effectue en trois phases :

1. **Initialisation** : Cette étape consiste à quantifier l'arbre de jonction avec les potentiels de croyance. Le résultat est un arbre de jonction inconsistant vu que les assignations initiales des potentiels ne satisfont pas la propriété de consistance locale formalisée par l'équation (3.9).
2. **Propagation globale** : Elle consiste en une série de passages de messages sur l'arbre de jonction. Ces passages de messages réarrangent les potentiels de l'arbre de jonction de telle sorte à aboutir à une structure consistante. Ainsi, le résultat de cette phase est un arbre de jonction consistant.
3. **Marginalisation** : Elle consiste à calculer, pour chaque variable d'intérêt V , la distribution de probabilités $P(A)$ à partir de l'arbre de jonction consistant. D'une manière détaillée, les étapes sont décrites comme suit :

Initialisation : La procédure suivante initialise les fonctions potentielles associées à l'arbre de jonction en utilisant les probabilités conditionnelles initiales :

Algorithme 3.5 : Initialisation Debut

Pour chaque cluster C_i **Faire** $\psi_{C_i}^I := 1$;
Pour chaque séparateur S_{ij} **Faire** $\psi_{S_{ij}}^I := 1$;
Pour chaque variable A **Faire**
 Sélectionner un cluster C_i contenant $\{A\} \cup U_A$;
 $\psi_{C_i}^I := \psi_{C_i}^I * P(A | U_A)$;
Fin

Les fonctions potentielles obtenues codifient la distribution probabiliste jointe associée à l'arbre de jonction, à savoir :

$$P = P_{\mathcal{J}\mathcal{T}} \quad (3.12)$$

où P est la distribution de possibilités du réseau causal initial obtenue en utilisant l'équation (3.1) et $P_{\mathcal{J}\mathcal{T}}$ représente la distribution de probabilités jointe associée à l'arbre de jonction obtenue en utilisant l'équation (3.8).

Après avoir initialisé les potentiels de l'arbre de jonction, il faudrait s'assurer de la consistance locale en effectuant une propagation globale décrite dans la section suivante :

La propagation globale : Elle induit à chaque cluster de passer des messages à l'ensemble de ses clusters adjacents. Ces messages sont ordonnés afin que chaque passage de message préserve la consistance introduite par le passage de message précédent. Une fois la propagation globale achevée, chaque paire (cluster, séparateur) est consistante et l'arbre de jonction est localement consistant.

Le mécanisme de passages des messages peut être explicité en détaillant la procédure de passage de message entre deux clusters adjacents, puis en explicitant comment se fait la coordination des passages de messages :

1. **Passage de messages simples** : Soient C_i et C_j deux clusters adjacents séparés par un séparateur S_{ij} et soient ψ_{C_i}, ψ_{C_j} et $\psi_{S_{ij}}$ leurs potentiels respectifs. Les différents potentiels sont mis à jour comme suit :

- (a) Préserver le même potentiel pour
- C_i
- :

$$\psi_{C_i}^{t+1} \leftarrow \psi_{C_i}^t \quad (3.13)$$

- (b) Assigner un nouveau potentiel à
- S_{ij}
- :

$$\psi_{S_{ij}}^{t+1} \leftarrow \sum_{C_i \setminus S_{ij}} \psi_{C_i}^t \quad (3.14)$$

- (c) Assigner un nouveau potentiel à
- C_j
- :

$$\psi_{C_j}^{t+1} \leftarrow \psi_{C_j}^t * \frac{\psi_{S_{ij}}^{t+1}}{\psi_{S_{ij}}^t} \quad (3.15)$$

2. **Coordination de messages multiples** : A partir d'un arbre de jonction de n clusters, l'algorithme de propagation globale sélectionne un cluster pivot arbitraire et effectue $2^*(n-1)$ passages de messages en deux phases :
- **Collecte de l'évidence** : Chaque cluster passe un message à ses clusters adjacents dans la direction du pivot, en commençant par des clusters les plus éloignés de celui-ci.
 - **Distribution de l'évidence** : Chaque cluster envoie des messages à ses clusters adjacents, en s'éloignant de la direction du pivot et ceci en commençant par le pivot lui-même jusqu'à atteindre les feuilles.

Remarque 3.5 *Dans le schéma de passages de messages, un cluster ne transmet un message à un voisin dans une direction, qu'après avoir reçu les messages de l'ensemble de ses nœuds (clusters) adjacents. Cette condition permet de garantir la consistance locale de l'arbre de jonction une fois la propagation globale achevée [89, 11].*

La procédure de propagation globale peut être résumée comme suit :

Algorithme 3.6 : Propagation globale

Début

Selection de la racine de la propagation

Soit Pivot, un cluster arbitraire C_i représentant la racine de la propagation ;

Soit Postorder, un vecteur contenant l'ordre dans lequel les messages vont être envoyés dans la phase de la collecte de l'évidence, ;

Soit Preorder, un vecteur contenant l'ordre dans lequel les messages vont être envoyés dans la phase de la distribution de l'évidence ;

Collecte de l'évidence

Pour $i := 1$ **jusqu'à** $\text{lenght}(\text{Postorder})-1$ **Faire**

$C_i \leftarrow \text{Postorder}[i]$;

$C_j \leftarrow$ cluster adjacent de C_i dans Postorder ;

Envoyer le message de C_i vers C_j en utilisant les équations (3.13), (3.14) et (3.15) ;

Distribution de l'évidence

Pour $i := 1$ **jusqu'à** $\text{lenght}(\text{Preorder})$ **Faire**

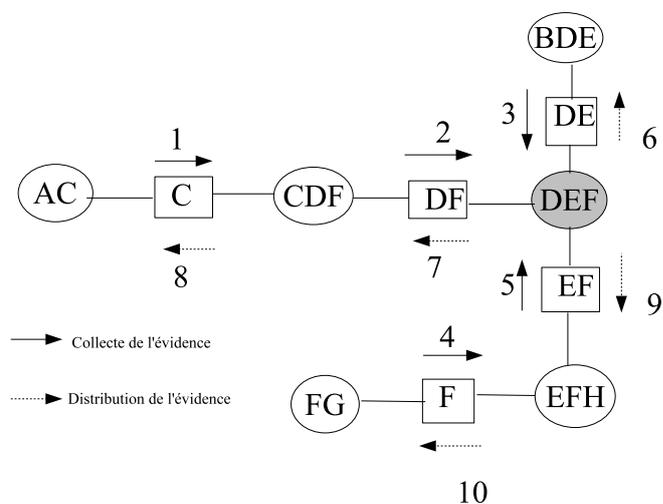


FIGURE 3.18 – Passage de messages durant la propagation globale associé à l'arbre de jonction de la figure 3.16

$C_i \leftarrow \text{Preorder}[i];$

$\text{Below} \leftarrow \text{cluster adjacents de } C_i \text{ dans Preorder};$

Pour $j := 1$ **jusqu'à** $\text{lenght}(\text{Below})$ **Faire**

$C_j \leftarrow \text{Below}[j];$

Envoyer le message de C_i à C_j en utilisant les équations (3.13), (3.14) et (3.15);

Fin

Exemple 3.10 La figure 3.18 illustre le passage de messages dans l'arbre de jonction représenté par la figure 3.16 lors de la procédure de la propagation globale.

Marginalisation : Une fois avoir obtenu un arbre de jonction consistant, la probabilité P associée à chaque variable d'intérêt A est calculée comme suit :

Algorithme 3.7 : Marginalisation

Debut

Identifier un cluster C_i contenant A ;

Calculer $P(A)$ en marginalisant $\psi_{C_i}^C$ sur A :

$$P(A) \leftarrow \sum_{C_i \setminus A} \psi_{C_i}^C;$$

Fin

Prise en compte d'une évidence : Les procédures présentées dans la section précédente peuvent être étendues dans le cas général où il s'agirait de calculer la probabilité d'une variable A , étant donnée l'observation d'une évidence totale e , notée $P(A | e)$.

La notion la plus simple de l'évidence est l'observation : Une observation est une expression de la forme $A=a$. Le *likelihood* associé à A , noté Λ_A , est le potentiel associé à

A, défini par :

$$\Lambda_A(a) = \begin{cases} 1 & \text{si A n'est pas instanciée} \\ 1 & \text{si A est instanciée à } a \\ 0 & \text{si A est instanciée mais pas à } a \end{cases} \quad (3.16)$$

La procédure de propagation pour les graphes à connexions multiples avec variables d'observation est identique à la procédure précédente (sans variable d'observation) à laquelle des étapes supplémentaires sont ajoutées. Elle se résume comme suit :

1. Initialisation :

L'étape d'initialisation est modifiée en introduisant une étape additionnelle qui consiste à initialiser la fonction *likelihood*, associée à chaque variable, à 1.

2. Prise en charge de l'observation :

La fonction *likelihood* est incorporée à l'arbre de jonction comme suit :

- (a) Encoder l'observation $A=a$ comme une fonction *likelihood* Λ_A^{new}
- (b) Identifier un cluster C_i contenant A
- (c) Mettre à jour ψ_{C_i} et Λ_A comme suit :

$$\begin{aligned} \psi_{C_i} &\leftarrow \psi_{C_i} * \Lambda_A^{new} \\ \Lambda_A &\leftarrow \Lambda_A^{new} \end{aligned}$$

En introduisant un ensemble d'observations, l'arbre de jonction encode $\mathcal{P}_{\mathcal{JT}}(A \wedge e)$ au lieu de $\mathcal{P}_{\mathcal{JT}}(A)$.

3. Normalisation :

La marginalisation d'un potentiel d'un cluster quelconque ψ_{C_i} dans une variable A permet d'obtenir la probabilité de A et e, définie par :

$$P(A, e) = \sum_{C_i \setminus A} \psi_{C_i}$$

Le but est de calculer $P(A | e)$, en normalisant $P(A \wedge e)$ comme suit :

$$P(A | e) = \frac{P(A \wedge e)}{P(e)} = \frac{P(A \wedge e)}{\sum_A P(A \wedge e)} \quad (3.17)$$

3.3 Réseaux causaux possibilistes

Les réseaux causaux possibilistes sont des contreparties possibilistes des réseaux causaux probabilistes. Ils sont issus de la théorie des possibilités. Contrairement à la logique possibiliste (issue également de la théorie des possibilités), il existe peu de travaux relatifs aux réseaux causaux possibilistes dans les domaines de la propagation et de l'apprentissage.

Pour la propagation, dans [56], les auteurs proposent une propagation possibiliste pour les hypergraphes. Dans [57], l'auteur présente une adaptation possibiliste de l'algorithme de Pearl. Pour les graphes possibilistes indirectes, Gebhardt, Kruse et Borgelt ont développé une boîte à outils, appelée POSSINFER [60, 59, 58]. Shenoy [52] propose une version possibiliste de la propagation utilisant les *Valuation-Based Systems*.

Dans le domaine de l'apprentissage des connaissances incertaines, dans [90], les auteurs proposent une méthode d'apprentissage à partir des connaissances incertaines.

En théorie des possibilités deux types de conditionnement ont été définis (voir section 1.5.3) basés sur deux opérateurs différents, (le produit et le minimum). Ceci a conduit à définir deux types de réseaux causaux possibilistes :

- les réseaux causaux possibilistes basés sur le produit qui sont proches du modèle graphique probabiliste, où les interprétations sont numériques,
- les réseaux causaux possibilistes basés sur le minimum, définis dans un contexte ordinal et qui posent le problème de cohérence dû au non-recouvrement des données initiales.

3.3.1 Définition d'un réseau causal possibiliste

Un réseau causal possibiliste, comme le cas d'un réseau causal probabiliste, est défini par :

- Une composante graphique : matérialisée par un Graphe Acyclique Indirect (DAG) \mathcal{G} représentant les relations de dépendances entre les variables.
- Une composante numérique : qui consiste à quantifier les différents liens représentés par le graphe en utilisant les distributions possibilistes conditionnelles de chaque nœud dans le contexte de ses parents. Ces dernières doivent obéir à la condition de normalisation.

Pour chaque variable A :

- Si $U_A = \emptyset$ (A est un nœud racine), alors les possibilités a priori relatives à la variable A doivent satisfaire :

$$\max_a \Pi(a) = 1, \forall a \in D_A,$$

- Si $U_A \neq \emptyset$, alors les distributions conditionnelles de la variable A dans le contexte de ses parents doivent satisfaire :

$$\max_a \Pi(a | a_A) = 1, \forall a \in D_A, u_A \in D_{U_A}$$

La différence de la définition du conditionnement a conduit à définir deux types de réseaux causaux possibilistes :

3.3.1.1 Les réseaux causaux possibilistes basés sur le produit :

Dans le contexte de la théorie des possibilités quantitative, un réseau causal possibiliste basé sur le produit, noté ΠG^* , est un graphe possibiliste où le conditionnement est basé sur le produit :

Définition 3.16 (Conditionnement possibiliste basé sur le produit)

$$\pi(\omega |_p \phi) = \begin{cases} \frac{\pi(\omega)}{\Pi(\phi)} & \text{si } \omega \models \phi \\ 0 & \text{sinon} \end{cases} \quad (3.18)$$

La distribution de possibilités jointe associée à un graphe causal possibiliste basé sur le produit est définie par le biais de la règle de chaînage basée sur le produit :

Définition 3.17 (Règle de chaînage basée sur le produit) *La distribution de possibilités globale jointe d'un réseau causal possibiliste basé sur le produit ΠG_p , à travers un ensemble de variables $V = \{A_1, A_2, \dots, A_N\}$ est exprimée comme le produit des N distributions a priori et des distributions conditionnelles à travers la règle de chaînage suivante :*

$$\pi_p(A_1, \dots, A_N) = \prod_{i=1..N} \Pi(A_i | U_{A_i}) \quad (3.19)$$

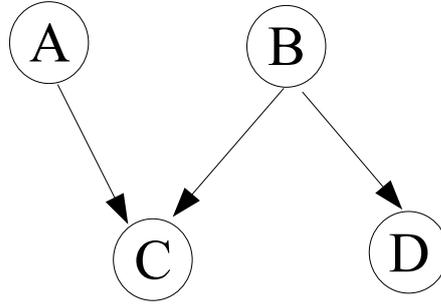


FIGURE 3.19 – Exemple de DAG

3.3.1.2 Les réseaux causaux possibilistes basés sur le minimum :

Dans le contexte de la théorie des possibilités qualitative, un réseau causal possibiliste basé sur le minimum, noté ΠG_m , est un graphe possibiliste où le conditionnement est basé sur le minimum :

$$\pi(\omega \mid_m \phi) = \begin{cases} 1 & \text{si } \pi(\omega) = \Pi(\phi) \text{ et } \omega \models \phi \\ \pi(\omega) & \text{si } \pi(\omega) < \Pi(\phi) \text{ et } \omega \models \phi \\ 0 & \text{sinon} \end{cases} \quad (3.20)$$

La distribution de possibilités jointe associée à un graphe causal possibiliste basé sur le minimum est définie par le biais de la règle de chaînage basée sur le minimum :

Définition 3.18 (Règle de chaînage basée sur le minimum) *La distribution de possibilités globale jointe d'un réseau causal possibiliste basé sur le minimum ΠG_m , à travers un ensemble de variables $V = \{A_1, A_2, \dots, A_N\}$ est exprimée comme le minimum des N distributions a priori et des distributions conditionnelles à travers la règle de chaînage suivante :*

$$\pi_m(A_1, \dots, A_N) = \min_{i=1..N} \Pi(A_i \mid U_{A_i}) \quad (3.21)$$

Exemple 3.11 *Considérons le graphe possibiliste basé sur le minimum de la figure (3.19). Les distributions de possibilités conditionnelles sont représentées dans la table 3.6. Les distributions de possibilités jointes, obtenues en utilisant la règle de chaînage basée sur le minimum, sont données par la table (3.7).*

A	$\Pi(A)$	B	$\Pi(B)$	D	B	$\Pi(D B)$
a	.5	b	1	d	b	1
$\neg a$	1	$\neg b$.4	d	$\neg b$	1
				$\neg d$	b	.3
				$\neg d$	$\neg b$.2

C	A	B	$\Pi(C A \wedge B)$	C	A	B	$\Pi(C A \wedge B)$
c	a	b	.6	$\neg c$	a	b	1
c	a	$\neg b$	1	$\neg c$	a	$\neg b$.8
c	$\neg a$	b	1	$\neg c$	$\neg a$	b	.3
c	$\neg a$	$\neg b$	0	$\neg c$	$\neg a$	$\neg b$	1

TABLE 3.7 – Distributions de possibilités jointes utilisant la règle de chaînage basée sur le minimum

A	B	C	D	$\pi_m(A \wedge B \wedge C \wedge D)$	A	B	C	D	$\pi_m(A \wedge B \wedge C \wedge D)$
a	b	c	d	.5	$\neg a$	b	c	d	1
a	b	c	$\neg d$.3	$\neg a$	b	c	$\neg d$.3
a	b	$\neg c$	d	.5	$\neg a$	b	$\neg c$	d	.3
a	b	$\neg c$	$\neg d$.3	$\neg a$	b	$\neg c$	$\neg d$.3
a	$\neg b$	c	d	.4	$\neg a$	$\neg b$	c	d	0
a	$\neg b$	c	$\neg d$.2	$\neg a$	$\neg b$	c	$\neg d$	0
a	$\neg b$	$\neg c$	d	.4	$\neg a$	$\neg b$	$\neg c$	d	.4
a	$\neg b$	$\neg c$	$\neg d$.2	$\neg a$	$\neg b$	$\neg c$	$\neg d$.2

Contrairement aux réseaux causaux basés sur le produit, les réseaux causaux basés sur le minimum ne respectent pas toujours la propriété de recouvrement des données initiales. Ceci influe sur la cohérence du réseau :

Définition 3.19 (Cohérence d'un réseau causal possibiliste basé sur le minimum)

Un réseau causal possibiliste basé sur le minimum ΠG_m est cohérent si l'application de la règle de chaînage correspondante à l'équation(3.21), permet de retrouver les distributions initiales fournies par l'expert : $\Pi_m(a | u_A) = \Pi(a | u_A)$

Ce qui n'est pas systématiquement le cas pour les réseaux causaux possibilistes basés sur le minimum. Néanmoins, il a été montré dans [61] que les données non recouvertes correspondent aux données redondantes, et de ce fait elles peuvent être ignorées sans qu'il n'y ait des répercussions sur la distribution globale jointe.

3.3.2 Propagation dans les réseaux causaux possibilistes basés sur le minimum

Dans ce qui suit, nous allons présenter les différents algorithmes de propagation dans le réseaux causaux possibilistes basés sur le minimum pour les polyarbres et les graphes à connexions multiples. Ces algorithmes, développés dans [61], sont des adaptations des algorithmes relatifs aux réseaux causaux bayésiens. Plus précisément, nous allons présenter dans cette section une version centralisée de l'algorithme de propagation de Pearl, tout en mettant en exergue le fait que, le problème de non-recouvrement des données initiales n'affecte pas le processus de propagation. Par la suite, une adaptation de l'algorithme de propagation de l'arbre de jonction probabiliste sera donnée.

3.3.2.1 Propagation basée sur le minimum dans les polyarbres

L'algorithme de propagation développé dans [61] est une adaptation de l'algorithme de propagation pour les polyarbre de la section 3.2.2.2. Cette adaptation est légèrement différente de celle proposée pour les réseaux causaux possibilistes basés sur le produit car des étapes supplémentaires sont nécessaires afin de transformer les distributions conditionnelles initiales en des distributions locales jointes.

Remarque 3.6 Dans ce qui suit, les mêmes notations que la section 3.2.2.2 vont être utilisées.

Principe de la propagation Le but de la propagation est de calculer, pour chaque nœud A, la distribution conditionnelle associée basée sur une totale évidence, notée $Bel_{Cdt}(a)$ et définie par : $Bel_{Cdt}(a) = \Pi_m(a | e)$.

Cette valeur est obtenue à partir de la distribution jointe, notée $Bel_{joint}(a)$ et définie par :

$$Bel_{joint}(a) = \Pi_m(a \wedge e) = \min(\lambda(a), \mu(a)) \quad (3.22)$$

où $\lambda(a) = \Pi_m(a \wedge \bar{e}_A)$ et $\mu(a) = \Pi_m(a \wedge e_A^+)$.

Ainsi, $\forall a \in D_A$, le calcul de $Bel_{Cdt}(a) = \Pi_m(a | e)$ peut être effectué en utilisant l'équation $Bel_{joint}(a) = \Pi_m(a \wedge e)$ et en appliquant la définition du conditionnement basée sur le minimum (équation 3.20) comme suit :

$$Bel_{Cdt}(a) = \Pi_m(a | e) = \begin{cases} \Pi_m(a \wedge e) & \text{si } \Pi_m(a \wedge e) < \Pi_m(e) = \max_{a \in D_A} \Pi_m(a \wedge e) \\ 1 & \text{sinon} \end{cases} \quad (3.23)$$

L'expression de Bel_{joint} dépend des valeurs λ et μ relatives au nœud A qui dépendent, respectivement, des λ -messages reçus des enfants (notés par $\lambda_{Y_j}(A)$ où λ_{Y_j} est le message reçu par A à partir du fils Y_j) et des μ -messages reçus des parents (notés par $\mu_A(U_i)$) où μ_A est le message que A reçoit de son parent U_i). Formellement, ces deux valeurs sont définies par :

– **Calcul de la valeur $\lambda \forall a \in D_A$,**

$$\lambda(a) = \Pi_m(a \wedge e_A^-) = \min(\lambda_A(a), \min_{j=1}^m \lambda_{Y_j}(a)) \quad (3.24)$$

où $\min_{j=1}^m \lambda_{Y_j}(a)$ correspond au minimum entre le λ -message reçu à partir du fils de A et $\lambda_A(a)$ dénote l'évidence locale de A, telle que :

$$\Lambda_A(a) = \begin{cases} 0 & \text{si A est instanciée à } e_A \neq a \\ 1 & \text{si A est instanciée à } e_A = a \text{ ou A n'est pas instanciée} \end{cases} \quad (3.25)$$

– **Calcul de la valeur $\mu \forall a \in D_A$,**

$$\mu(a) = \Pi_m(a \wedge e_A^+) = \max_u \min(\Pi(a \wedge u), \min_{i=1}^n \mu_A(u_i)) \quad (3.26)$$

où $\min_{i=1}^n$ correspond au minimum entre les μ -messages reçus à partir des parents de A.

– **Calcul de λ -message** Le message λ de A à ses parents $U_i, (i \in \{1, \dots, n\})$, lorsque $U_i = u_i$ est défini par :

$$\lambda_A(u_i) = \Pi_m(e_{U_i A}^- \wedge u_i) = \max_{a \in D_A} \min[\lambda(a), \max_{u_k: k \neq i} (\min(\Pi(a \wedge u), \min_{k \neq i} \mu_A(u_k)))] \quad (3.27)$$

– **Calcul de μ -message** Le message μ de A à ses enfants $Y_j, (j \in \{1, \dots, n\})$ lorsque $A=a$ est défini par :

$$\mu_{Y_j}(a) = \Pi_m(a \wedge e_{AY_j}^+) = \min(\lambda_A(a), \min_{i=1 \dots n, i \neq j} \lambda_{Y_i}(a), \mu(a)) \quad (3.28)$$

Algorithme de propagation : Le principe de la propagation dans un réseau causal possibiliste basé sur le minimum est illustré par l'algorithme suivant :

Algorithme 3.12 : Propagation basée sur le minimum dans un polyarbre

Début

Choix de la racine de propagation

- Soit \mathcal{S} le plus petit ensemble connecté incluant les nœuds observés ;
- Soit pivot un nœud arbitraire dans \mathcal{S} représentant la racine de la propagation ;
- Soit Postorder un vecteur contenant l'ordre dans lequel les messages sont envoyés dans la phase de collecte de l'évidence (le dernier nœud étant le pivot) ;
- Soit Preorder un vecteur contenant l'ordre dans lequel les messages sont envoyés dans la phase de collecte de l'évidence (le premier nœud est le pivot)

Initialisation

- Initialiser l'ensemble des valeurs λ et μ messages à 1 ;
- Pour chaque nœud A racine, la valeur $\mu(a)$ est initialisée à $\Pi(a)$, $\forall a \in D_A$.
- Pour chaque nœud A observé, assigner à $\lambda(a)$ la valeur 1 si A est instanciée à a , et lui assigner la valeur 0 dans le cas contraire.

Collecte de l'évidence

Pour $i := 1$ à $\text{length}(\text{Postorder}) - 1$ **Faire**

- $A \leftarrow \text{Postorder}[i]$;
- $B \leftarrow$ un nœud adjacent de A dans Postorder ;
- Calculer $\lambda(A)$ en utilisant l'équation (3.24) ;
- Calculer $\mu(A)$ en utilisant l'équation (3.26) ; **Si** B est un nœud parent de A **alors**

Envoyer un message λ de A vers B en utilisant l'équation (3.27)

Sinon

Envoyer un message μ de A à B en utilisant l'équation (3.28)

Distribution de l'évidence

Pour $i := 1$ à $\text{length}(\text{Preorder}) - 1$ **Faire**

- $A \leftarrow \text{Preorder}[i]$;
- $\text{Below} \leftarrow$ un nœud adjacent de A dans Preorder ;
- Calculer $\lambda(A)$ en utilisant l'équation (3.24) ;
- Calculer $\mu(A)$ en utilisant l'équation (3.26) ;

Pour $j := 1$ à $\text{length}(\text{Below})$ **Faire**

- $B \leftarrow \text{Below}[j]$;

- **Si** B est un nœud parent de A **alors**

Envoyer un message λ de A vers B en utilisant l'équation (3.27) **Sinon**

Envoyer un message μ de A à B en utilisant l'équation (3.28)

Marginalisation Pour chaque nœud A , $\text{Bel}(A) = \Pi_p(a \mid e)$ est calculé en utilisant l'équation (3.22) ;

Normalisation

- Calculer $\text{Bel}_{\text{Cdt}}(A)$ à partir de $\text{Bel}_{\text{Joint}}(A)$ en utilisant l'équation (3.23) ;

Fin

3.3.2.2 Propagation basée sur le minimum dans les graphes à connexions multiples

La méthode de propagation dans ce type de DAG est similaire à celle présentée dans le cas probabiliste, qui consiste d'abord à transformer le réseau initial en un arbre de jonction puis de procéder à la propagation sur la nouvelle structure.

A partir d'un arbre de jonction \mathcal{JT} , l'unique distribution de possibilités globale jointe est définie par :

$$\pi_{\mathcal{JT}}(A_1, \dots, A_N) = \min_{i=1 \dots N} \pi_{C_i} \quad (3.29)$$

où N représente le nombre de clusters dans \mathcal{JT} . Dans le cas où l'arbre de jonction est consistant, le potentiel attaché à chaque cluster C_i doit satisfaire :

$$\pi_{C_i} = \Pi_m(C_i) \quad (3.30)$$

En utilisant cette dernière équation, la distribution de possibilités associée à une variable A dans un arbre de jonction consistant est calculée en utilisant un cluster C_i contenant la variable A , en marginalisant son potentiel sur A comme suit :

$$\Pi_m(A) = \max_{C_i \setminus A} \pi_{C_i} \quad (3.31)$$

Après la phase de transformation graphique, la propagation consiste à calculer le degré de possibilité de n'importe quelle variable $A \in V$, $\Pi_m(A)$. Elle s'effectue en trois phases :

1. **Initialisation** : Cette étape consiste à quantifier l'arbre de jonction, en utilisant les probabilités conditionnelles initiales :

Algorithme 3.13 : Initialisation

Debut

Pour chaque cluster C_i **Faire** $\pi_{C_i}^I \leftarrow 1$;

Pour chaque séparateur S_{ij} **Faire** $\pi_{S_{ij}}^I \leftarrow 1$;

Pour chaque variable A **Faire**

 Sélectionner un cluster C_i contenant $\{A\} \cup U_A$;

$\pi_{C_i}^I := \min(\pi_{C_i}^I, \Pi(A | U_A))$;

Fin

L'arbre de jonction, ainsi initialisé, codifie les mêmes distributions que le graphe initial. Ainsi,

$$\pi_m = \pi_{\mathcal{JT}}^I \quad (3.32)$$

où π_m est la distribution de possibilités jointe du réseau causal initial obtenue en utilisant l'équation (3.21) et $\pi_{\mathcal{JT}}^I$ représente la distribution de probabilités jointe associée à l'arbre de jonction obtenue en utilisant l'équation (3.29).

Après avoir initialisé les potentiels de l'arbre de jonction, il faudrait s'assurer de la consistance locale en effectuant la propagation globale décrite ci-après.

2. **La propagation globale** : Elle est réalisée par le mécanisme de passages de messages entre chaque cluster C_i et ses clusters adjacents C_j , comme dans le cas probabiliste.

Si un nœud C_i envoie un message à ses clusters adjacents C_j , les potentiels de C_i , C_j et leurs séparateurs seront mis à jour comme suit :

- (a) Préserver le même potentiel pour C_i :

$$\pi_{C_i}^{t+1} \leftarrow \pi_{C_i}^t \quad (3.33)$$

- (b) Affecter un nouveau potentiel à S_{ij} :

$$\pi_{S_{ij}}^{t+1} \leftarrow \max_{C_i \setminus S_{ij}} \pi_{C_i}^t \quad (3.34)$$

(c) Affecter un nouveau potentiel à C_j :

$$\pi_{C_j}^{t+1} \leftarrow \min(\pi_{C_j}^t, \pi_{S_{i_j}}^{t+1}) \quad (3.35)$$

La procédure de propagation globale peut être résumée comme suit [61] :

Algorithme 3.14 : Propagation globale

Debut

Selection de la racine de la propagation :

Soit Pivot un cluster arbitraire C_i représentant la racine de la propagation ;

Soit Postorder un vecteur contenant l'ordre dans lequel les messages vont être envoyés dans la phase de la collecte de l'évidence ;

Soit Preorder un vecteur contenant l'ordre dans lequel les messages vont être envoyés dans la phase de la distribution de l'évidence ;

Collecte de l'évidence :

Pour $i := 1$ jusqu'à $\text{lenght}(\text{Postorder})-1$ Faire

$C_i \leftarrow \text{Postorder}[i]$;

$C_j \leftarrow$ cluster adjacent de C_i dans Postorder ;

Envoyer le message de C_i vers C_j en utilisant les équations (3.33), (3.34) et (3.35) ;

Distribution de l'évidence :

Pour $i := 1$ jusqu'à $\text{lenght}(\text{Preorder})$ Faire

$C_i \leftarrow \text{Preorder}[i]$;

$\text{Below} \leftarrow$ cluster adjacents de C_i dans Preorder ;

Pour $j := 1$ jusqu'à $\text{lenght}(\text{Below})$ Faire

$C_j \leftarrow \text{Below}[j]$;

Envoyer le message de C_i à C_j en utilisant les équations (3.33), (3.34) et (3.35) ;

Fin

Remarque 3.7 A partir de ces deux procédures, il est important de relever certains points :

- A chaque étape t de la procédure de la propagation globale, l'arbre de jonction codifie la même distribution jointe. Formellement, $\forall t$, de l'étape initiale jusqu'à l'étape finale :

$$\pi_{\mathcal{JT}}^t = \pi_{\mathcal{JT}}^{t+1} \quad (3.36)$$

où $\pi_{\mathcal{JT}}^t$ représente la distribution jointe relative à l'arbre de jonction \mathcal{JT} de l'étape t .

- A partir de équations (3.21) et (3.29), nous déduisons que de la phase d'initialisation jusqu'à la phase de consistance globale, l'arbre de jonction codifie la même distribution jointe. Formellement,

$$\pi_m = \pi_{\mathcal{JT}}^C \quad (3.37)$$

où π_m est la distribution jointe de ΠG_m et $\pi_{\mathcal{JT}}^C$ est la distribution jointe associée à \mathcal{JT} après la procédure de propagation globale.

- La consistance globale est atteinte après les phases de collecte de l'évidence et de distribution de l'évidence.

Marginalisation : A partir de l'arbre de jonction consistant obtenu lors de la phase précédente, la mesure de possibilités associée à chaque variable d'intérêt A , notée $\Pi_m(A)$ peut être calculée par l'algorithme suivant :

Algorithme 3.15 : Marginalisation

Début

Identifier un cluster C_i contenant A ;

Calculer $\Pi_m(A)$ en marginalisant $\pi_{C_i}^C$ sur A :

$$\Pi_m(A) \leftarrow \max_{C_i \setminus A} \pi_{C_i}^C ;$$

Fin

Prise en charge de l'évidence : La prise en charge d'une évidence totale e se fait par la généralisation des différents algorithmes précédents. Il s'agit de calculer $\forall A \in V, \Pi_m(A \wedge e)$. L'évidence peut être codifiée en utilisant la fonction *likelihood* définie par :

$$\Lambda_A(a) = \begin{cases} 1 & \text{si } A \text{ n'est pas instanciée} \\ 1 & \text{si } A \text{ est instanciée à } a \\ 0 & \text{si } A \text{ est instanciée mais pas à } a \end{cases} \quad (3.38)$$

La prise en compte de l'évidence e nécessite l'extension de la procédure d'initialisation afin d'incorporer la connaissance certaine. En effet, l'évidence e doit être codifiée comme une fonction *likelihood*, en utilisant l'équation (3.38), puis elle doit être intégrée dans l'arbre de jonction en ajoutant des étapes supplémentaires à la procédure d'initialisation : Elle se résume comme suit :

- Pour chaque instance de la variable A , codifier l'observation $A=a$ comme une fonction *likelihood*, en utilisant l'équation (3.38).
- Identifier un cluster C_i contenant A :

$$\pi_{C_i}^I \leftarrow \min(\pi_{C_i}^I, \Lambda_A)$$

La marginalisation d'un potentiel d'un cluster quelconque $\pi_{C_i}^C$ dans une variable A , en utilisant l'équation (3.31) permet d'obtenir la mesure de possibilité de A et de e :

$$\Pi_m(A \wedge e) = \max_{C_i \setminus A} \pi_{C_i}^C$$

Le calcul de $\Pi_m(A | e)$ est effectué à partir de $\Pi_m(A \wedge e)$ en appliquant la définition du conditionnement basé sur le produit comme suit :

$$\Pi_m(A | e) = \begin{cases} \Pi_m(A \wedge e) & \text{si } \Pi_m(A \wedge e) < \Pi_m(e) = \max_A \Pi_m(A \wedge e) \\ 1 & \text{sinon} \end{cases} \quad (3.39)$$

3.4 Réseaux causaux basés sur le produit

Un réseau causal possibiliste basé sur le produit, noté ΠG_p , est un graphe possibiliste où le conditionnement est basé sur le produit :

$$\pi(\omega |_p \phi) = \begin{cases} \frac{\pi(\omega)}{\Pi(\phi)} & \text{si } \omega \models \phi \\ 0 & \text{sinon} \end{cases} \quad (3.40)$$

La distribution de possibilités jointe associée à un graphe causal possibiliste basé sur le produit est définie par le biais de la règle de chaînage basée sur le produit :

Définition 3.20 (Règle de chaînage basée sur le produit) La distribution de possibilités globale jointe d'un réseau causal possibiliste basé sur le produit ΠG_p , à travers un ensemble de variables $V=\{A_1, A_2, \dots, A_N\}$ est exprimée comme le produit des N distributions a priori et distributions conditionnelles à travers la règle de chaînage suivante :

$$\pi_p(A_1, \dots, A_N) = \prod_{i=1..N} \Pi(A_i | U_{A_i}) \quad (3.41)$$

Exemple 3.12 Considérons le graphe possibiliste basé sur le produit de la figure 7.1. Les distributions de possibilités conditionnelles sont représentées par la table 7.1. Les distributions de possibilités jointes, obtenues en utilisant la règle de chaînage, sont données par la table 7.2.

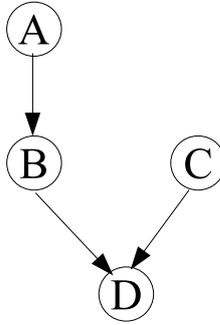


FIGURE 3.20 – Exemple d'un DAG

TABLE 3.8 – Distributions initiales						
A	$\Pi(A)$	C	$\Pi(C)$	B	A	$\Pi(B A)$
a	.2	c	1	b	a	.6
$\neg a$	1	$\neg c$.7	b	$\neg a$	1
				$\neg b$	a	1
				$\neg b$	$\neg a$.1

D	B	C	$\Pi(D B \wedge C)$	D	B	C	$\Pi(D B \wedge C)$
d	b	c	.9	$\neg d$	b	c	1
d	b	$\neg c$.8	$\neg d$	b	$\neg c$	1
d	$\neg b$	c	1	$\neg d$	$\neg b$	c	.5
d	$\neg b$	$\neg c$	1	$\neg d$	$\neg b$	$\neg c$.3

Les réseaux causaux possibilistes basés sur le produit vérifient deux propriétés primordiales :

1. **Critère de Normalisation** : La marginalisation produit toujours des distributions normalisées. D'une manière formelle, soit X un sous ensemble de V et x une instantiation de X . Soit $Z=V/X$. alors,

$$\max_z \left\{ \prod \{ \Pi(a | u_A) : a \in z, u_A \subseteq z \wedge x \} \right\} = 1 \quad (3.42)$$

2. **Recouvrement des données initiales**. Il signifie que les distributions de possibilités initiales prodiguées par l'expert sont retrouvées à partir de la distribution de possibilités jointe donnée par l'équation (7.2).

TABLE 3.9 – Distributions de possibilités jointes utilisant la règle de chaînage basée sur le produit

A	B	C	D	$\pi_p(A \wedge B \wedge C \wedge D)$	A	B	C	D	$\pi_p(A \wedge B \wedge C \wedge D)$
a	b	c	d	.108	$\neg a$	b	c	d	.90
a	b	c	$\neg d$.012	$\neg a$	b	c	$\neg d$	1
a	b	$\neg c$	d	.0672	$\neg a$	b	$\neg c$	d	.56
a	b	$\neg c$	$\neg d$.084	$\neg a$	b	$\neg c$	$\neg d$.70
a	$\neg b$	c	d	.02	$\neg a$	$\neg b$	c	d	.1
a	$\neg b$	c	$\neg d$.1	$\neg a$	$\neg b$	c	$\neg d$.05
a	$\neg b$	$\neg c$	d	.14	$\neg a$	$\neg b$	$\neg c$	d	.07
a	$\neg b$	$\neg c$	$\neg d$.42	$\neg a$	$\neg b$	$\neg c$	$\neg d$.21

D'une manière formelle, pour un nœud A quelconque d'un graphe possibiliste basé sur le produit noté ΠG_p , la distribution de possibilités conditionnelle $\Pi_p(a | u_A)$ calculée à partir de la distribution de possibilités globale jointe π_p , est égale à la distribution conditionnelle donnée par l'expert concernant le nœud A, notée $\Pi(a | u_A)$. Ainsi,

$$\Pi_p(a | u_a) = \Pi(a | u_A) \quad (3.43)$$

Exemple 3.13 Soit le DAG précédent. A titre d'exemple, l'égalité suivante est vérifiée : $\Pi_p(d | b \wedge \neg c) = \Pi(d | b \wedge \neg c) = .8$, vu que $\Pi(d | b \wedge \neg c) = \frac{\Pi_p(d \wedge b \wedge \neg c)}{\Pi_p(b \wedge \neg c)} = \frac{.56}{.7} = .8$.

3.5 Propagation basée sur le produit dans les polyarbres

L'algorithme présenté dans cette section est une adaptation *possibiliste* de l'algorithme de propagation présenté dans la section 3.2.2.2. Il est basé sur les communications locales via deux types de messages, λ -messages et μ -messages, circulant, respectivement des nœuds enfants vers les nœuds parents et des nœuds parents vers les nœuds enfants.

Remarque 3.8 La formulation mathématique de l'algorithme de propagation pour les réseaux causaux possibilistes dans le cas d'un polyarbre est identique à celle correspondant aux réseaux causaux bayésiens, avec une différence qui consiste à utiliser l'opérateur maximum au lieu de l'opérateur addition.

Ainsi,

- $\forall a \in D_A$, la mesure de probabilité courante de a basée sur l'évidence totale e est définie par :

$$Bel(a) = \Pi_p(a | e) = \alpha * \lambda(a) * \mu(a) \quad (3.44)$$

où $\alpha = \frac{1}{\max_a Bel(a)}$ est le facteur de normalisation

- La valeur λ associée à chaque instance $a \in D_a$ est définie par :

$$\lambda(a) = \Pi_p(a | e_A^-) = \lambda_A(a) * \prod_{j=1}^m \lambda_{Y_j}(a) \quad (3.45)$$

où $\lambda_A(a)$ dénote l'évidence totale relative au nœud A telle que :

$$\lambda_A(a) = \begin{cases} 0 & \text{si } e_A \neq a \text{ (A est instanciée à } a \text{ avec } e_A \neq a) \\ 1 & \text{sinon (A est instanciée à } a \text{ avec } e_A = a \text{ ou A n'est pas instanciée)} \end{cases} \quad (3.46)$$

- La valeur μ associée à chaque instance $a \in D_A$ est définie par :

$$\mu(a) = \Pi_p(a | e_A^+) = \max_u \Pi(a | u) * \prod_{i=1}^n \mu_A(u_i) \quad (3.47)$$

- Le message λ de A vers ses parents U_i , ($i \in \{1, \dots, n\}$) où $U_i = u_i$ est défini par :

$$\lambda_A(u_i) = \Pi_p(e_{U_i A}^- | u_i) = \beta * \max_a \lambda(a) [\max_{u_k: k \neq i} \Pi(a | u) * \prod_{k \neq i} \mu_A(u_k)] \quad (3.48)$$

où β est une constante de normalisation.

Remarque 3.9 Dans le cas où chaque nœud dans le graphe a un parent unique alors la valeur λ est simplifiée comme suit :

$$\lambda_A(u_i) = \beta * \max_a \lambda(a) * \Pi(a | u)$$

- Le message μ de A à ses enfants Y_j , ($j \in \{1, \dots, m\}$) où $A = a$ est défini par :

$$\mu_{Y_j}(a) = \Pi_p(a | e_A^+ Y_j) = \alpha * \lambda_A(a) * \prod_{i=1 \dots m, i \neq j} \lambda_{Y_i}(a) * \mu(a) \quad (3.49)$$

3.5.1 Algorithme de propagation :

L'algorithme de propagation s'effectue essentiellement en deux grandes étapes :

- La première phase consiste à sélectionner arbitrairement un nœud pivot dans le plus petit ensemble des nœuds connectés contenant l'ensemble des nœuds modifiés (noté \mathcal{S})
- La deuxième phase consiste à faire passer les messages en deux étapes :
 - *Collecte de l'évidence* où chaque nœud dans \mathcal{S} fait passer un message à ses nœuds adjacents dans la direction du pivot, en commençant par le nœud le plus éloigné du nœud pivot.
 - *Distribution de l'évidence* où chaque nœud du graphe fait passer un message à ses nœuds adjacents, en commençant par le pivot lui-même jusqu'à atteindre les feuilles du DAG.

Les grandes lignes de cet algorithme sont les suivantes :

Algorithme 3.8 : Propagation basée sur le produit dans un polyarbre

Début

Choix de la racine de propagation

- Soit \mathcal{S} le plus petit ensemble connecté incluant les nœuds observés ;
- Soit pivot un nœud arbitraire dans \mathcal{S} représentant la racine de la propagation ;
- Soit Postorder un vecteur contenant l'ordre dans lequel les messages sont envoyés dans la phase de collecte de l'évidence (le dernier nœud étant le pivot) ;
- Soit Preorder un vecteur contenant l'ordre dans lequel les messages sont envoyés dans la phase de distribution de l'évidence (le premier nœud est le pivot)

Initialisation

- Initialiser l'ensembles des valeurs λ , μ et messages à 1 ;
- Pour chaque nœud racine A, la valeur $\mu(a)$ est initialisée à $\Pi(a)$, $\forall a \in D_A$.
- Pour chaque nœud A observé, il faut assigner à $\lambda(a)$ la valeur 1 si A est instanciée à a, et lui assigner la valeur 0 dans le cas contraire.

Collecte de l'évidence

Pour $i := 1$ à $\text{length}(\text{Postorder}) - 1$ **Faire**

- $A \leftarrow \text{Postorder}[i]$;
- $B \leftarrow$ un nœud adjacent de A dans Postorder ;
- Calculer $\lambda(A)$ en utilisant l'équation (7.6);
- Calculer $\mu(A)$ en utilisant l'équation (7.8); **Si** B est un nœud parent de A **alors**
Envoyer un message λ de A vers B en utilisant l'équation (7.9)

Sinon

Envoyer un message μ de A à B en utilisant l'équation (7.10)

Distribution de l'évidence

Pour $i := 1$ à $\text{length}(\text{Preorder}) - 1$ **Faire**

- $A \leftarrow \text{Preorder}[i]$;
- $\text{Below} \leftarrow$ un nœud adjacent de A dans Preorder ;
- Calculer $\lambda(A)$ en utilisant l'équation (7.6);
- Calculer $\mu(A)$ en utilisant l'équation (7.8);

Pour $j := 1$ à $\text{length}(\text{Below})$ **Faire**

- $B \leftarrow \text{Below}[j]$;
- **Si** B est un nœud parent de A **alors**
Envoyer un message λ de A vers B en utilisant l'équation (7.9) **Sinon**
Envoyer un message μ de A à B en utilisant l'équation (7.10)

Marginalisation Pour chaque nœud A , $\text{Bel}(A) = \Pi_p(a | e)$ est calculé en utilisant l'équation (7.5);

Fin

3.6 Propagation basée sur le produit dans les réseaux causaux possibilistes à connexions multiples

Pour ce type de DAG, le principe de la propagation est similaire à celui de la propagation dans le cadre des réseaux bayésiens. Il consiste à transformer le DAG initial en un arbre de jonction, puis d'effectuer la propagation sur ce dernier.

L'arbre de jonction, noté \mathcal{JT} , est défini et construit de manière similaire que dans le cas probabiliste (voir section 3.2.2.3).

Pour chaque cluster C_i (respectivement séparateur S_{ij}) de \mathcal{JT} , des distributions jointes locales (ou des potentiels) sont assignées, notées π_{C_i} (respectivement $\pi_{S_{ij}}$).

A partir de \mathcal{JT} , une distribution de possibilités jointe globale unique, notée $\pi_{\mathcal{JT}}$, est assignée. Elle est définie par :

$$\pi_{\mathcal{JT}}(A_1, \dots, A_N) = \frac{\prod_{i=1}^m \pi_{C_i}}{\prod_{j=1}^{m-1} \pi_{S_{ij}}} \quad (3.50)$$

où m est le nombre de clusters dans le \mathcal{JT} .

Un arbre de jonction est caractérisé par la propriété de consistance suivante :

Définition 3.21 Soient C_i et C_j deux clusters adjacents dans un arbre de jonction \mathcal{JT} et soit S_{ij} leur séparateur. Un lien entre C_i et C_j est dit stable ou consistant si

$$\max_{C_i \setminus S_{ij}} \pi_{C_i} = \pi_{S_{ij}} = \max_{C_j \setminus S_{ij}} \pi_{C_j} \quad (3.51)$$

où $\max_{C_i \setminus S_{ij}} \pi_{C_i}$ est la distribution marginale de S_{ij} définie à partir de π_{C_i} .

Si l'ensemble des liens d'un arbre de jonction est consistant, alors l'arbre de jonction est dit globalement consistant.

Lorsque l'arbre de jonction est globalement consistant, alors le potentiel attaché à chaque cluster correspond à sa distribution jointe locale calculée à partir du graphe initial. Formellement,

$$\pi_{C_i} = \Pi_p(C_i) \quad (3.52)$$

A partir de cette dernière équation, la distribution de possibilités d'une variable $A \in V$, dans un arbre de jonction globalement consistant est calculée en utilisant un cluster quelconque C_i contenant A en marginalisant son potentiel sur A comme suit :

$$\Pi_p(A) = \max_{C_i \ni A} \pi_{C_i} \quad (3.53)$$

Dans ce qui suit, $\pi_{C_i}^I$ représente le potentiel du cluster C_i lors de l'étape d'initialisation et $\pi_{C_i}^C$ représente le potentiel du cluster C_i lors de l'étape de consistance globale.

3.6.1 Algorithme de la propagation

Après avoir construit l'arbre de jonction optimal associé au graphe causal original, l'étape suivante consiste à calculer la composante numérique associée à l'arbre de jonction. En d'autres termes, pour chaque variable $A \in V$, il faudrait calculer le degré de possibilité $\Pi(A)$. Le problème de la prise en charge d'une évidence totale e (calculer $\Pi(a | e)$) sera abordé par la suite.

Après la phase de transformation graphique, la propagation s'effectue en trois phases :

1. **Initialisation** : Cette étape consiste à quantifier l'arbre de jonction, en utilisant les probabilités conditionnelles initiales : **Algorithme 3.9 : Initialisation**

Debut

Pour chaque cluster C_i **Faire** $\pi_{C_i}^I \leftarrow 1$;
Pour chaque séparateur S_{ij} **Faire** $\pi_{S_{ij}}^I \leftarrow 1$;
Pour chaque variable A **Faire**

Sélectionner un cluster C_i contenant $\{A\} \cup U_A$;

$\pi_{C_i}^I := \pi_{C_i}^I * P(A | U_A)$;

Fin

L'arbre de jonction ainsi initialisé codifie les mêmes distributions que le graphe initial :

$$\pi_p = \pi_{\mathcal{JT}}^I \quad (3.54)$$

où π_p est la distribution de possibilités jointe du réseau causal initial obtenue en utilisant l'équation (7.2) et $\pi_{\mathcal{JT}}^I$ représente la distribution de probabilités jointe associée à l'arbre de jonction obtenue en utilisant (7.11). Après avoir initialisé les potentiels de l'arbre de jonction, il faudrait s'assurer de la consistance locale en effectuant la propagation globale décrite comme suit :

2. **La propagation globale** : La propagation globale est réalisée par le mécanisme de passages de messages entre chaque cluster C_i et ses clusters adjacents C_j en deux phases, en commençant par un cluster arbitraire dit nœud pivot :

– **Collecte de l'évidence :**

Chaque cluster passe un message à ses clusters adjacents dans la direction du pivot, en commençant par les clusters les plus éloignés de celui-ci. L'ordre dans lequel les messages sont envoyés est matérialisé par *Postorder*.

– **Distribution de l'évidence :**

Chaque cluster envoie des messages à ses clusters adjacents, en commençant par le pivot lui même jusqu'à atteindre les feuilles. L'ordre dans lequel les messages sont envoyés est matérialisé par *Preorder*.

Pour les deux phases, lorsqu'un cluster C_i fait passer un message à ses nœuds adjacents C_j , les potentiels associés à C_i, C_j et leur séparateur S_{ij} sont mis à jour comme suit :

(a) Preserver le même potentiel pour C_i :

$$\pi_{C_i}^{t+1} \leftarrow \pi_{C_i}^t \quad (3.55)$$

(b) Assigner un nouveau potentiel à S_{ij} :

$$\pi_{S_{ij}}^{t+1} \leftarrow \max_{C_i \setminus S_{ij}} \pi_{C_i}^t \quad (3.56)$$

(c) Assigner un nouveau potentiel à C_j :

$$\pi_{C_j}^{t+1} \leftarrow \pi_{C_j}^t * \frac{\pi_{S_{ij}}^{t+1}}{\pi_{S_{ij}}^t} \quad (3.57)$$

La procédure de propagation globale peut être résumée comme suit [61] : **Algorithme**

3.10 : Propagation globale

Debut

Selection de la racine de la propagation

Soit Pivot un cluster arbitraire C_i représentant la racine de la propagation ;

Soit Postorder un vecteur contenant l'ordre dans lequel les messages vont être envoyés dans la phase de la collecte de l'évidence ;

Soit Preorder un vecteur contenant l'ordre dans lequel les messages vont être envoyés dans la phase de la distribution de l'évidence ;

Collecte de l'évidence

Pour $i := 1$ jusqu'à $\text{lenght}(\text{Postorder})-1$ Faire

$C_i \leftarrow \text{Postorder}[i];$

$C_j \leftarrow$ cluster adjacent de C_i dans *Postorder* ;

Envoyer le message de C_i vers C_j en utilisant les équations (7.16), (7.17) et (7.18) ;

Distribution de l'évidence

Pour $i := 1$ jusqu'à $\text{lenght}(\text{Preorder})$ Faire

$C_i \leftarrow \text{Preorder}[i];$

Below \leftarrow cluster adjacents de C_i dans *Preorder* ;

Pour $j := 1$ jusqu'à *length* (*Below*) **Faire**

$C_j \leftarrow$ *Below*[j];

Envoyer le message de C_i à C_j en utilisant les équations (7.16), (7.17) et (7.18);

Fin

Remarque 3.10 A partir de ces deux procédures, il est important de relever certains points :

- A chaque étape t de la procédure de la propagation globale, l'arbre de jonction codifie la même distribution jointe. Formellement,

$$\pi_{\mathcal{JT}}^t = \pi_{\mathcal{JT}}^{t+1} \quad (3.58)$$

où $\pi_{\mathcal{JT}}^t$ représente la distribution jointe relative à l'arbre de jonction \mathcal{JT} à l'étape t .

- A partir des équations (7.15) et (7.19), nous déduisons que de la phase d'initialisation jusqu'à la phase de consistance globale, l'arbre de jonction codifie la même distribution jointe. Formellement,

$$\pi_p = \pi_{\mathcal{JT}}^C \quad (3.59)$$

où π_p est la distribution jointe de ΠG_p et $\pi_{\mathcal{JT}}^C$ est la distribution jointe associée à \mathcal{JT} après la procédure de propagation globale.

- La consistance globale est atteinte après les phases de collecte de l'évidence et de distribution de l'évidence.

3. **Marginalisation** A partir de l'arbre de jonction consistant obtenu lors de la phase précédente, la mesure de possibilités associée à chaque variable d'intérêt A , notée $\Pi_p(A)$ peut être calculée par l'algorithme suivant :

Algorithme 3.11 : Marginalisation

Debut

Identifier un cluster C_i contenant A ;

Calculer $\Pi_p(A)$ en marginalisant $\pi_{C_i}^C$ sur A :

$$\Pi_p(A) \leftarrow \max_{C_i \setminus A} \pi_{C_i}^C ;$$

Fin

Prise en charge de l'évidence : La prise en charge d'une évidence totale e se fait par la généralisations des différents algorithmes précédents. Il s'agit de calculer $\forall A \in V, \Pi_p(A \wedge e)$. L'évidence peut être codifiée en utilisant la fonction *likelihood* définie par :

$$\Lambda_A(a) = \begin{cases} 1 & \text{si } A \text{ n'est pas instanciée} \\ 1 & \text{si } A \text{ est instanciée à } a \\ 0 & \text{si } A \text{ est instanciée mais pas à } A \end{cases} \quad (3.60)$$

La prise en charge de l'évidence e nécessite l'extension de la procédure d'initialisation afin d'incorporer la connaissance certaine. En effet, l'évidence e doit être codifiée comme

une fonction *likelihood*, en utilisant l'équation (7.21), puis elle doit être intégrée dans l'arbre de jonction en ajoutant des étapes supplémentaires à la procédure d'initialisation. Elle se résume comme suit :

- Pour chaque instance de la variable A, codifier l'observation A=a comme une fonction *likelihood*, en utilisant l'équation (7.21).
- Identifier un cluster C_i contenant A :

$$\pi_{C_i}^I \leftarrow \pi_{C_i}^I * \Lambda_A$$

La marginalisation d'un potentiel d'un cluster quelconque $\pi_{C_i}^C$ dans une variable A, en utilisant l'équation (7.14) permet d'obtenir la mesure de possibilité de A et e :

$$\Pi_p(A, e) = \max_{C_i \setminus A} \pi_{C_i}^C$$

Le calcul de $\Pi_p(A | e)$ est effectué à partir de $\Pi_p(A \wedge e)$ en appliquant la définition du conditionnement basé sur le produit comme suit :

$$\Pi_p(A | e) = \frac{\Pi_p(A \wedge e)}{\Pi_p(e)} = \frac{\Pi_p(A \wedge e)}{\max_A \Pi_p(A \wedge e)} \quad (3.61)$$

Remarque 3.11 *Nous rappelons que la tâche de trouver une triangulation optimale est un problème NP-complet [78].*

3.7 Conclusion

Le traitement de l'incertitude est une issue fondamentale de l'intelligence artificielle. L'approche la plus utilisée de nos jours est sans aucun doute l'approche bayésienne. L'avantage de celle-ci est qu'elle bénéficie des techniques de la théorie des probabilités. De plus, du point de vue philosophique, un agent *rationnel* affecte des incertitudes à des événements de façon à satisfaire les axiomes de probabilités. Elle offre une représentation naturelle des connaissances par des liens entre causes et effets. La représentation des connaissances est graphique ce qui induit sa lisibilité.

Néanmoins, l'approche bayésienne a été largement critiquée car des probabilités subjectives doivent être affectées à chaque événement.

Du point de vue du raisonnement, les différents algorithmes de propagation développés sont complexes dans le cas des réseaux causaux à connexions multiples dans le cadre de la théorie de probabilités ou dans le cadre de la théorie des possibilités. Ces derniers ont été obtenus en adaptant les algorithmes développés dans le contexte possibiliste qui offre un cadre qualitatif et un cadre quantitatif.

Chapitre 4

L'inférence en logique possibiliste

4.1 Introduction

Nous rappelons que la logique des possibilités, issue de la logique floue, a été conçue pour raisonner automatiquement sur des connaissances incertaines. Une base de connaissance possibiliste est composée d'un ensemble de couples (ϕ, α) , où ϕ est une formule logique classique $\in \mathcal{L}$ et α représente le seuil minimal du degré de nécessité de ϕ . Autrement dit : $N(\phi) \geq \alpha$.

Cette valuation peut être interprétée différemment selon le contexte. En effet,

- Dans un contexte qualitatif, l'attribution d'une valuation α à une formule ϕ permet d'établir un ordre implicite entre les connaissances. Cette interprétation des degrés de certitude définit *la logique possibiliste standard*.
- Dans un contexte numérique ou quantitatif, la valeur α est interprétée comme une valuation numérique de la connaissance ϕ quantifiant l'incertitude qui lui est affiliée définissant ainsi la notion de *logique possibiliste quantitative* qui est souvent comparée à la logique probabiliste.

Vu cette distinction, les mécanismes d'inférence mis en œuvre sont certainement différents.

4.2 L'inférence en logique possibiliste standard (SPL)

En logique possibiliste standard, les bases de connaissances sont stratifiées et la strate la plus élevée correspond aux connaissances les plus certaines. Ainsi, le mécanisme de raisonnement consiste, à partir d'une SPL Σ , à dériver des conclusions en utilisant la partie la plus certaine de celle-ci.

Le problème de déduction en logique possibiliste standard est basé sur le principe d'ordre induit par les connaissances [32]. Il peut être résumé comme suit :

Soit Σ une base possibiliste standard et ϕ une formule logique classique que l'on veut déduire de Σ avec un certain degré. Le processus d'inférence revient à calculer la plus grande valeur de α telle que (ϕ, α) est une conséquence logique de Σ . Pour aborder les différents mécanismes de raisonnement en SPL, il est nécessaire de présenter certains concepts.

- **La satisfaction d'une base de connaissances par une distribution de possibilités :** Une distribution de possibilités π satisfait une formule SPL (ϕ, α) (noté $\pi \models (\phi, \alpha)$) ssi $N(\phi) \geq \alpha$, où N est la mesure de nécessité induite par π .

Par extension, une distribution de possibilités π satisfait une base de connaissances SPL $\Sigma = \{(\phi_i, \alpha_i), i = 1..n\}$ (noté $\pi \models \Sigma$) ssi $\forall i, \pi \models (\phi_i, \alpha_i)$

- **La conséquence logique :** Une formule (ϕ, α) est une conséquence logique d'une base de connaissances SPL Σ ssi toute distribution de possibilités π qui satisfait Σ , satisfait alors (ϕ, α) .

Le problème de déduction est alors présenté comme suit :

Soit Σ une base de connaissances SPL et ϕ une formule logique classique à déduire à partir de Σ avec un certain degré. Il faudrait alors calculer le plus grand degré α tel que (ϕ, α) est une conséquence logique de Σ . En d'autres termes, il faut calculer :

$$Val(\phi, \Sigma) = \sup\{\alpha \in [0, 1] : \Sigma \models (\phi, \alpha)\}$$

- **Le principe du minimum de spécificité :** Un résultat fondamental qui découle de la déduction à partir d'une base de connaissances SPL est qu'il existe toujours une distribution de possibilités la moins spécifique qui satisfait la base de connaissances Σ .

Formellement, si $\Sigma = \{(\phi_i, \alpha_i), i = 1..n\}$, alors la distribution de possibilités la moins spécifique π_Σ qui satisfait Σ est définie par :

$$\pi_\Sigma(\omega) = \begin{cases} 1 & \text{si } \omega \models \phi_1 \wedge \dots \wedge \phi_n \\ \min\{1 - \alpha_i : \omega \models \neg\phi_i, (\phi_i, \alpha_i) \in \Sigma\} & \text{sinon} \end{cases} \quad (4.1)$$

Ainsi, pour toute distribution de possibilités [34] π , π satisfait Σ ssi $\pi \geq \pi_\Sigma$ D'où :

$$\Sigma \models (\phi, \alpha) \text{ssi } \pi_\Sigma \models (\phi, \alpha)$$

L'incohérence partielle : Un des objectifs majeurs de la logique possibiliste standard est qu'elle permet de considérer l'incohérence partielle et donc de permettre la déduction à partir d'une base de connaissances partiellement incohérente.

Une base de connaissances possibiliste Σ est partiellement incohérente si sa distribution possibiliste est telle que :

$$0 < \sup \pi_\Sigma < 1$$

Mesurer le degré d'incohérence de Σ revient à évaluer à quel degré il existe au moins une interprétation complètement possible pour Σ . Le degré d'incohérence d'une base de connaissances possibiliste est défini par :

$$Incons(\Sigma) = 1 - \sup_{\omega \in \Omega} \pi_\Sigma(\omega)$$

- **La déduction sous l'incohérence partielle :** Le processus de déduction en SPL est proche du processus en logique classique. Le lien entre le degré d'incohérence en SPL et l'incohérence d'une base classique est défini par :

$$Incons(\Sigma) = \sup\{\alpha : \Sigma_\alpha^* \text{ est incohérente}\} = \inf\{\alpha : \Sigma_\alpha^* \text{ est consistante}\}$$

où $\Sigma = \{(\phi_i, \alpha_i) : i = 1..n\}$ est une base possibiliste standard,

$\Sigma_\alpha^* = \{\phi : (\phi, \beta) \in \Sigma, \beta \geq \alpha\}$ est la projection classique de la α -cut de Σ ,

$\Sigma_\alpha^* = \{\phi : (\phi, \beta) \in \Sigma, \beta > \alpha\}$ est la projection classique de la α -cut stricte de Σ .

Les généralisations des théorèmes de déduction et de réfutation de la logique classique en SPL sont représentées par les résultats suivants [34] :

– La déduction :

$$\Sigma \cup \{(\phi, 1)\} \models (\psi, \alpha) \text{ssi} (\phi \rightarrow \psi, \alpha)$$

– La réfutation :

$$\Sigma \models (\phi, \alpha) \text{ssi} \Sigma \cup \{(\neg\phi, 1)\} \models (\perp, \alpha)$$

qui est équivalent à :

$$\text{Val}(\phi, \Sigma) = \text{Incons}(\Sigma \cup \{(\neg\phi, 1)\})$$

Ce résultat montre que tout problème de déduction en logique possibiliste revient à un calcul d'un degré d'incohérence.

L'équation suivante montre qu'afin de déduire (ϕ, α) , seules les formules dont les poids sont supérieurs ou égaux à α sont utilisées. Ainsi,

$$\Sigma \models (\phi, \alpha) \text{ssi} \Sigma_\alpha \models (\phi, \alpha)$$

Ainsi, ces résultats permettent de calculer $\text{Val}(\phi, \Sigma)$ par dichotomie, où Σ est une SPL et ϕ est une formule propositionnelle à déduire et ceci en utilisant n'importe quel prouveur pour le problème de satisfiabilité propositionnelle SAT.

Soit $\alpha_0 = 0$ et soient $\alpha_1, \dots, \alpha_n$ les valuations distinctes qui apparaissent dans Σ , ordonnées d'une manière croissante : $0 < \alpha_1, \dots, \alpha_n \leq 1$.

La procédure de calcul de $\text{Val}(\phi, \Sigma)$ peut être résumée comme suit :

Algorithme 8.1 : Calcul de $\text{Val}(\phi, \Sigma)$

Debut

$l := 0;$

$u := n;$

Tant que $(l < u)$ **Faire**

$r := \lfloor (l+u)/2 \rfloor;$

Si $(\Sigma_{\alpha_r}^* \wedge \neg\phi)$ **est consistante** **Alors**

$u := r-1;$

sinon

$l := r;$

$\text{Val}(\phi, \Sigma) := \alpha_r;$

Fin

Cet algorithme contient exactement $\lceil \log_2 n \rceil$ appels à SAT pour le test de satisfiabilité, où n représente le nombre des différents strates qui existent dans Σ [32].

Le problème de déduction en logique possibiliste standard revient donc à calculer un degré d'incohérence. L'algorithme proposé par Lang utilise la stratification des bases de connaissances. Les formules des strates les moins prioritaires étant les moins certaines sont rejetées plus facilement en cas d'incohérence.

Dans cette approche ordinale, les formules de faible priorité sont systématiquement rejetées dans le cas où elles sont en contradiction avec une formule donnée. Mais cette philosophie peut s'avérer inadéquate pour certains domaines. En effet, il est parfois souhaitable de faire intervenir la notion de compensation qui permet de décider s'il faudrait rejeter une seule formule de forte priorité ou rejeter un très grand nombre de formules de faible priorité en fonction du nombre et de l'importance des formules.

4.3 L'inférence en logique possibiliste quantitative

La majorité des algorithmes d'inférence se basent sur des tests de satisfiabilité utilisant des prouveurs de la famille SAT. Le problème SAT fût le premier problème prouvé comme étant NP-complet [91].

4.3.1 Le problème WMAXSAT

Le problème de satisfiabilité, SAT en abrégé, constitue l'épine dorsale de la discipline de la NP-complétude. Malgré le nombre important d'études effectuées dans ce domaine, il demeure pour les chercheurs, une de leurs importantes préoccupations. SAT est défini par les composantes suivantes :

- Soit $X = \{x_1, x_2, \dots, x_n\}$, un ensemble de n variables booléennes.
- Soit $C = \{c_1, c_2, \dots, c_m\}$, un ensemble de m clauses où :
 - chaque clause est une disjonction de littéraux,
 - chaque littéral est une variable (un littéral positif) ou sa négation (un littéral négatif).
- Soit D , la donnée SAT, composée par une conjonction de littéraux.

Le problème SAT consiste à déterminer s'il existe une assignation des variables x_i de X telle que la donnée D soit satisfaite (évaluée à VRAI). SAT est un problème de décision qui admet une réponse 'Oui' s'il existe une assignation de variables qui satisfait toutes les clauses ou 'Non' si aucune assignation n'existe.

Néanmoins, si la réponse est négative, une question importante en découle : quel est le nombre maximal de clauses pouvant être satisfaites simultanément ? A partir de cette question, est défini le problème du "Maximum Satisfiability", MAXSAT en abrégé. Le problème MAXSAT est un problème d'optimisation dont la fonction objective est le nombre de clauses satisfaites. Le problème MAXSAT est un problème NP-difficile.

Si de plus, nous associons à chaque clause c_i un poids $w(c_i)$, l'objectif revient donc à trouver une assignation aux variables x_i de X qui maximise la somme des poids des clauses satisfaites simultanément ou encore qui minimise la somme des poids des clauses insatisfaites, d'où la définition du problème MAX-SAT pondéré ou WMAXSAT (Weighted MAX-SAT).

L'algorithme d'inférence développé dans [92] exploite la procédure de transformation d'une base possibiliste quantitative et une base de pénalités. Dans une base possibiliste quantitative, à chaque formule est associée le seuil minimal de son degré de nécessité. Pour une base de pénalités, à chaque formule est associée une pondération qui correspond au prix à payer pour pouvoir la rejeter.

4.3.2 Principe de l'algorithme

Le principe de la déduction se déroule en trois étapes :

1. **Étape 1** : Elle consiste à déduire la base de pénalités PK associée à la base possibiliste quantitative Σ donnée. Soit

$$\Sigma = \{(\phi_i, \alpha_i), i = 1 \dots n\} \text{ avec } \alpha_i \in [0, 1]$$

La base de pénalités PK correspondante est définie par :

$$PK = \{(\phi_i, k_i), i = 1 \dots n\} \text{ avec } k_i \in \mathcal{N} + \infty \text{ où } k_i = -\log(1 - \alpha_i).$$

L'équivalence entre deux bases de connaissances est généralement prouvée à partir de la relation d'équivalence entre leurs distributions induites.

En effet :

$$\forall \omega, \pi_{\Sigma_{PK}}(\omega) = e^{-\kappa_{PK}(\omega)}$$

où $\pi_{\Sigma_{PK}}$ représente la distribution de possibilité associée à la base possibiliste quantitative Σ_{PK} et κ_{PK} représente la distribution associée à la base de pénalités PK. D'autre part, la logique des pénalités définit les deux mesures suivantes :

- **Le coût d'une interprétation ω** : Il est noté par $\kappa_{PK}(\omega)$ et il est égal à la somme des pénalités des formules dans PK violées par ω . D'une manière formelle, il est défini par :

$$\kappa_{PK}(\omega) = \begin{cases} 0 & \text{si } \forall(\phi_i, k_i) \in PK, \omega \models \phi_i \\ \sum \kappa_i \{ \omega \models \neg \phi_i, (\phi_i, k_i) \in PK \} & \text{sinon} \end{cases} \quad (4.2)$$

- **Le coût d'une formule ϕ** : Il est noté par $K_{PK}(\phi)$ et il correspond au coût à payer afin de rendre la formule ϕ consistante avec PK. D'une manière formelle, il est défini par :

$$K_{PK}(\phi) = \min(\kappa_{PK}(\omega) : \omega \models \phi) \quad (4.3)$$

2. **Étape 2** : la philosophie d'une base de pénalités réside dans le fait que les poids sont des pénalités à payer afin de violer une formule. Ainsi, une formule est d'autant plus importante que son poids est élevé. Une formule inviolable (donc certaine) a un poids infini.

Dans cette optique, la recherche d'un sous ensemble consistant de formules est réalisée par le paramètre qui correspond à la somme minimale des pénalités des formules qui doivent être rejetées afin de restaurer la cohérence de la base. D'où l'utilisation d'un prouveur de type WMAXSAT (Weighted MAXSAT).

Un prouveur WMAXSAT prend comme paramètre d'entrée une base pondérée et en sortie, il fournit :

- l'interprétation ω qui minimise la somme des coûts des formules qu'elle falsifie
- la somme minimale des coûts des formules falsifiées par ω .

Le processus de fonctionnement de WMAXSAT peut être résumé comme suit :

Procédure WMAXSAT($PK, \omega, k_{PK}(\omega)$)

Debut

Entrée : la base de pénalités PK

Sorties : $\omega, k_{PK}(\omega)$ (somme minimale des coûts des formules falsifiées par ω)

Répéter

choisir une interprétation ω' de Ω

$k' := 0$;

sommecouts $:= +\infty$;

Pour chaque formule ϕ_i in PK **Faire**

Si $\omega' \not\models \phi_i$ **Alors**

$k' := k' + \alpha_i$;

Si $k' <$ sommecouts **Alors**

sommecouts $:= k'$;

$\omega := \omega'$;

jusqu'à(sommecouts=0) ou (il n'existe plus d'interprétation)

$k_{PK}(\omega) :=$ sommecouts ;

retourner($\omega, k_{PK}(\omega)$) ;

Fin Donc le problème d'inférence est redéfini comme suit :

- soit $PK = \{(\phi_i, k_i) \mid i = 1, \dots, n \text{ et } k_i \in \mathcal{N} \cup \infty\}$,
- soit E une évidence, à laquelle est associée un coût infini, qui illustre le fait qu'elle est inviolable,
- soit ψ la variable d'intérêt à inférer.

Il s'agit alors de tester si $PK \cup (E, \infty) \models (\psi, \beta)$, en déduisant son coût β .

D'une manière formelle, WMAXSAT fournit une interprétation ω' qui minimise la somme des coûts des formules falsifiées dans $PK \cup \{(Evidence, +\infty) \cup (\psi, +\infty)\}$ ainsi que ce coût $k_{PK}(\omega')$. Deux cas peuvent se présenter :

- si $k_{PK}(\omega') = +\infty$, alors $PK \cup \{(E, \infty)\} \not\models \psi$.
- si $k_{PK}(\omega') \neq +\infty$, alors $\omega' \models \{(E, +\infty)\}$ et $\omega' \models \{(\psi, +\infty)\}$.

Dans le deuxième cas, nous avons :

$$K_{PK}(\psi) = \min(k_{PK}(\omega) : \omega \models \psi) = k_{PK}(\omega') \quad (4.4)$$

- L'attribution d'un coût infini à l'évidence E reflète le fait qu'elle est inviolable. Donc si une interprétation ω viole l'évidence, son coût serait alors infini.
- Pour la variable d'intérêt ψ , nous lui assignons également un coût infini pour signifier qu'elle doit être également satisfaite.

Ainsi, l'interprétation fournie par le prouveur WMAXSAT doit satisfaire $(E, +\infty)$ et $(\psi, +\infty)$.

3. **Étape 3 :** Dans le cas où $PK \cup (E, +\infty) \models (\psi, K_{PK}(\psi))$, il faudrait calculer alors le degré de possibilité de ψ .

Le calcul du degré de possibilité d'une formule ψ , noté $\Pi(\psi)$ à partir du coût de pénalité de cette formule est donné par :

$$\Pi(\psi) = e^{-K_{PK}(\psi)} \quad (4.5)$$

Ainsi, l'algorithme d'inférence est :

Procédure *infquant()*

Debut

genere(sigma,evidence,interet);

PK :=transformer(sigma);

infini :=calculinfini;

insérer(PK,evidence,infini);

insérer(PK,interet,infini);

WMAXSAT(PK,interpretation,coutinter);

Si (*coutinter=infini*) **Alors**

écrire("sigma n'infère pas la variable d'intérêt");

sinon

coutinteret :=coutinter;

degreinteret :=e^{-coutinteret};

écrire ("sigma infère",interet,"avec un degré",degreinteret);

Fin

Exemple 4.1 Soit à dérouler la procédure d'inférence sur la base possibiliste quantitative suivante :

$\Sigma = \{(\neg a, 0.999088), (b \vee a, 0.632121), (\neg b \vee \neg a, 0.997521), (c \vee a \vee \neg b, 0.864665), (\neg c \vee \neg a \vee b, 0.981684), (\neg c \vee \neg a \vee \neg b, 0.999088), (d \vee c, 0.999088), (\neg d \vee \neg c, 0.981684), (e \vee a \vee b \vee d, 0.981684), (e \vee a \vee \neg b \vee d, 0.999088), (e \vee a \vee b \vee \neg d, 0.999088), (e \vee a \vee \neg b \vee \neg d, 0.997521), (\neg e \vee \neg a \vee b \vee d, 0.999088), (\neg e \vee \neg a \vee \neg b \vee d, 0.950213), (\neg e \vee \neg a \vee b \vee \neg d, 0.999877), (\neg e \vee \neg a \vee \neg b \vee \neg d, 0.999665)\}$.

Soient *d* la variable d'évidence et $\neg c$ la variable d'intérêt.

- **Étape 1** : Elle consiste à déduire la base de pénalités *PK* équivalente à Σ .

$PK = \{(\neg a, 7), (b \vee a, 1), (\neg b \vee \neg a, 6), (c \vee a \vee \neg b, 2), (\neg c \vee \neg a \vee b, 4), (\neg c \vee \neg a \vee \neg b, 7), (d \vee c, 7), (\neg d \vee \neg c, 4), (e \vee a \vee b \vee d, 4), (e \vee a \vee \neg b \vee d, 7), (e \vee a \vee b \vee \neg d, 7), (e \vee a \vee \neg b \vee \neg d, 6), (\neg e \vee \neg a \vee b \vee d, 7), (\neg e \vee \neg a \vee \neg b \vee d, 3), (\neg e \vee \neg a \vee b \vee \neg d, 9), (\neg e \vee \neg a \vee \neg b \vee \neg d, 8)\}$

- **Étape 2** : Elle consiste à exécuter *PEN-WMAXSAT* ayant comme arguments d'entrée la base de pénalités *PK*, l'évidence *d* et la variable d'intérêt $\neg c$.

L'appel à la procédure *PEN-WMAXSAT* nous fournit une interprétation ω' telle que la somme des coûts des formules falsifiées par cette interprétation est minimale :

$\omega' = \neg a \vee \neg b \vee \neg c \vee d \vee f$ et $k_{PK}(\omega') = K_{PK}(\text{interet}) = 1$.

- **Étape 3** : Elle consiste à déduire le degré de possibilité de la variable d'intérêt à partir de son coût de pénalité en utilisant l'équation 8.4 :

$\Pi(\text{interet}) = e^{-K_{PK}(\text{interet})} = e^{-1} = .367879$

4.3.3 Complexité de l'algorithme d'inférence en logique des pénalités

Il est très attrayant d'utiliser la logique des pénalités comme mode de représentation des connaissances incertaines, vu qu'elle présente des propriétés intéressantes pour

l'inférence. En effet, la complexité du problème de déduction pour une base de connaissances à pénalités est une des plus simples parmi les modes d'inférence. Elle est de l'ordre de (Δ_2^P) [31].

La classe (Δ_2^P) est une classe de problèmes pouvant être résolus en un temps polynomial par une machine rehaussée avec un NP oracle. La classe NP est un sous ensemble de la classe (Δ_2^P) .

Cette complexité est considérée comme la plus simple [93]. Cependant, elle peut être excessive dans le cadre d'applications pratiques [31].

Bibliographie

- [1] D. Dubois and H. Prade. Uncertainty issues in artificial intelligence. In *Workshop on Philosophy of Probability*, 1990.
- [2] D.Kayser. *La représentation des connaissances*. Editions Hermès, 1997.
- [3] P.Smets. Varieties of ignorance and the need for well-founded theories. 1999.
- [4] P.Smets. Imperfect information : Imprecision- uncertainty. *UMIS*, 1999.
- [5] B.Bouchon-Meunier. *La logique floue et ses applications*. Addison-Wesley France, SA, 1995.
- [6] G. Shafer. A mathematical theory of evidence. In *New Jersey : Princeton University Press*, 1976.
- [7] L. Zadeh. Fuzzy sets. *Information and Control*, 8 :338–353, 1965.
- [8] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.
- [9] D. Dubois and H. Prade. Necessity measures and the resolution principle. *IEEE Trans. on Systems, Man and Cybernetics*, 17 :474–478, 1987.
- [10] J. Pearl. Probabilistic reasoning in intelligent systems : Networks of plausible inference. Morgan Kaufmann Publ. Inc., San Mateo, Ca, 1988.
- [11] F. V. Jensen. An introduction to Bayesian Networks. University College London : UCL Press, 1996.
- [12] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). In *J Royal Statist Soc B*, volume 50, pages 157–224, 1988.
- [13] L.A.Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information science*, 9 :43–80, 1975.
- [14] P.Fabiani. *Représentation dynamique de l'incertain et stratégies de prise d'information pour un système autonome en environnement évolutif*. PhD thesis, Ecole Nationale Supérieure de l'Aéronotique et de l'Espace, 1996.
- [15] D.Dubois and H.Prade. Representation and combinaison of uncertainty with belief functions and possibility mesures. In *Computer Intelligence 4*, pages 244–264, 1988.
- [16] D.Dubois and H.Prade. *Introduction aus systèmes experts de gestion*, chapter Le traitement de l'imprécision et de l'incertitude dans les modèles de raisonnement des experts. Eyrolles, 1985.
- [17] P.Smets. What is dempster-shafer model ? In J.Kacprzyk R.Yager and M.Fedrizzi, editors, *Advances in Dempster-Shafer theory of evidence*, pages 5–34, 1994.
- [18] P.Smets and P.Magrez. La représentation de la croyance. *Revue de l'intelligence artificielle*, pages 31–46, 1987.

- [19] P.Smets, E.H.Mamdani, D.Dubois, and H.prade. *Non-Standard Logics for Automated Reasoning*. 1988.
- [20] E.Ouzounova. *Incertitude et mesure de performance*. PhD thesis, Université de Fribourg Suisse, 2005.
- [21] A.P.Dempster. Upper and lower probabilities induced by a multivalued mapping. In *Ann Math Stat 38*, pages 325–339, 1967.
- [22] S.Parsons and E.H.Mamdani. Qualitative dempster-shafer theory. In *Decision Technologies*. Barcelone, June 1993.
- [23] N.Wilson. *Some theoretical aspects of the Dempster-Shafer theory*. PhD thesis, Departement of computing and mathematical sciences Oxford Polytechnic, 1992.
- [24] P. Smets. Belief Functions. In *Non-Standard Logics for Approximative Reasoning (P. Smets, A. Mamdani, D. Dubois and H. Prade eds)*, pages 253–286. New York : Academic Press, 1988.
- [25] P. Smets. The transferable belief model for quantified belief representation. In *Handbooks of Defeasible Reasoning and Uncertainty Management Systems, Vol. 1*, pages 267–301. (D. Gabbay and P. Smets, eds.) Kluwer, Dordrecht, 1998.
- [26] D.Dubois and H.Prade. Possibility theory and its applications : A retrospective and prospective view. In *IEEE International Conference on Fuzzy Systems*, pages 3–11, 2003.
- [27] D. Dubois and H. Prade. (with the collaboration of H. Farreny, R. Martin-Clouaire and C. Testemale). Possibility Theory - an Approach to Computerized Processing of Uncertainty. 1988.
- [28] D.Dubois and H.Prade. An introduction to possibilistic and fuzzy logics. In *Non-Standard Logics for Approximative Reasoning (P. Smets, A. Mamdani, D. Dubois and H. Prade eds)*, pages 287–326. New York : Academic Press, 1988.
- [29] E. Hisdal. Conditional possibilities independence and non interaction. *Fuzzy Sets and Systems*, 1, 1978.
- [30] Allet Hadj Ali. *Raisonnement qualitatif sur les ordres de grandeur relatif : Formalisation et traitement en logique floue*. Thèse de doctorat d'état, Université Mouloud Mameri Tizi Ouzou, 2004.
- [31] F. Bannay-Dupin de Saint-Cyr. Gestion de l'évolutif et de l'incertain en logiques pondérées. In *Thèse de doctorat. Université Paul Sabatier. Toulouse.*, 1996.
- [32] J. Lang. Possibilistic logic : Complexity and algorithms. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 5 :179–220, 2000.
- [33] S.Benferhat. *Evolution des croyances et raisonnement plausible - Modèles qualitatifs*. Thèse d'habilitation à diriger les recherches, Université Paul Sabatier Toulouse, 2000.
- [34] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, (D. Gabbay et al., eds, 3, Oxford University Press :pages 439–513, 1994.
- [35] G. Pinkas. Propositional nonmonotonic reasoning and inconsistency in symmetric neural networks. In *12th IJCAI, Sydney, Australia, Morgan-Kaufmann*, pages 525–530, 1991.
- [36] D. Dubois and H. Prade. Possibility theory as a basis for qualitative decision theory. In *14th International Joint Conference on Artificial Intelligence (IJCAI'95), Montréal*, pages 1924–1930, 1995.

- [37] S.Benferhat and F. Bannay-Dupin de Saint-Cyr. Contextual handling of conditional knowledge. In *5th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU96*, pages 1369–1374, 1996.
- [38] M.Freund. Default extension : an alternative to the probabilistic approach. 1993.
- [39] W.Sphon. A general non probabilistic theory of inductive reasoning. In *In proceeding of the fourth conference of uncertainty in Artificial Intelligence (UAI88)*, pages 149–158, 1988.
- [40] D. Dubois and H. Prade. Epistemic entrenchment and possibilistic logic. *Artificial Intelligence*, 50 :223–239, 1991.
- [41] R.Carnap. Logical foundations of probability. In *Routledge and Kegan Paul, London*, 1950.
- [42] J.Y.Halpern. An analysis of first-order logics of probability. In *In proceeding of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 1375–1381, August 20-25 1989.
- [43] L.Farinas del Cerro, A.Hezig, and J.Lang. Ordering-based non monotonic reasoning. *Artificial Intelligence*, (66) :375–393, 1994.
- [44] E.Rich. Default reasoning as likelihood reasoning. In *American Association for Artificial Intelligence Conference*, pages 348–351. Washington, DC, 1983.
- [45] H.Farreny and H.Prade. Default and inexact reasoning with possibility degrees. In *IEEE Trans System Man Cyber 16*, pages 270–276, 1986.
- [46] F.Dupin de Saint-Cyr, J.Lang, and T.Schiex. Gestion de l'inconsistance dans les bases de connaissances ; une approche syntaxique basée sur la logique des pénalités. In *9ème congrés Reconnaissance des Formes et Intelligence Artificielle*. Paris, France, 1994.
- [47] H.Raiffa. *Decision Analysis*. Addison-Welsley Publishing Compagny Toronto, 1968.
- [48] R.A.Howard and J.E.Matheson. Influence diagramms. *The principles and applications of decision Analysis*, 2 :720–761, 1984.
- [49] R.D.Shachter. Evaluating influence diagramms. *Operations Research*, 34 :871–882, 1986.
- [50] R.D.Shachter. A graph-based inferences method for conditional independance. In *Seventh Conference on Uncertainty in Artificial Intelligence UAI91*, pages 353–360. Morgan Kaufmann Publishers, 1991.
- [51] P.P.Shenoy. A fusion algorithm for solving bayesian decision problem. In *Seventh Conference on Uncertainty in Artificial Intelligence UAI91*, pages 169–198. Morgan Kaufmann Publishers, 1991.
- [52] P.P.Shenoy. Using possibility theory in expert systems. *Fuzzy Sets and Systems*, 52 :129–142, 1992.
- [53] P.P.Shenoy. Valuation-based systems for bayesian decision analysis. *Operations research*, 40 :463–484, 1992.
- [54] P.P.Shenoy. A comparaison of graphical technical techniques for decision analysis. *European Journal of Operational Reasearch*, 78 :1–21, 1994.
- [55] P.P.Shenoy. Conditional independance in valuations-based systems. *International Journal of Approximate Reasoning*, 10 :203–234, 1994.
- [56] D. Dubois and H. Prade. Inference in probabilistic hypergraph. In *International Conference on Information Processing of Uncertainty in Knowledge Bases Systems (IPMU'90)*, pages 228–230. Paris France, 1990.

- [57] P.Fonck. *Réseaux d'inférence pour le raisonnement possibiliste*. PhD thesis, Université de Liège, Faculté des sciences, 1994.
- [58] J. Gebhardt C.Borgelt and R. Kruse. Inference methods. In *Handbook of Fuzzy Computation*, pages 1–13. Institute of Physics Publishing,Bristol,United Kingdom, 1998.
- [59] J. Gebhardt and R. Kruse. Background and perspectives of possibilistic graphical models. In *4th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'97), LNAI 2143*, pages 108–121, 1997.
- [60] J. Gebhardt and R. Kruse. Possinfer- a software tool for possibilistic inference. In *Fuzzy Set Methods in Information Engineering : A guide tour of Applications*. D.Dubois, H.Prade, R.Yager,J.Wiley and sons, New York, 1995.
- [61] N.Ben Amor. *Qualitative possibilistic graphical models :from independance to propagation algorithms*. PhD thesis, ISG Université de Tunis, 2002.
- [62] J.Cano, M.Delgado, and S.Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *International Journal of Approximate Reasoning*, 8 :253–380, 1993.
- [63] F. V. Jensen. *Bayesian networks and Decision graphs*. Springer-Verlag, New York, 2001.
- [64] Mc Carthy and P.J.Hayes.
- [65] V.Isham. An introduction to spatial point processes and markov random fields. *International Statistical Review*, 49 :21–43, 1981.
- [66] D.R.Cox and N.Wermuth. *Multivariate Dependencies-Models, Analysis and Interpretation*. Chapman Hall, London, 1996.
- [67] S.L.Lauritzen. *Graphical Models*. Clarendon, Oxford UK, 1996.
- [68] S.L.Lauritzen. *Lectures on contengency tables*. University of Aalborg Press,Aalborg, Denmark, 1982.
- [69] S.L.Lauritzen and N.Wermuth. Graphical and recursive models for contingency tables. *biometrika*, 70 :37–52, 1983.
- [70] J.B.Carlin, H.Kiiveri, and T.P.Speed. Recursive causal models. *Journal of the Australian Mathematical Society*, 36 :30–52, 1984.
- [71] T.Verma. Causal networks : Semantics and expressiveness. In Inc Elsevier Publishing Compagny, editor, *Fourth Conference on Uncertainty in Artificial Intelligence (UAI'88)*, pages 69–76, New York, 1988.
- [72] D.Geiger, J.Pearl, and T.Verma. D-separation : from theorems to algorithms. In Inc Elsevier Publishing Compagny, editor, *Fifth Conference on Uncertainty in Artificial Intelligence (UAI'89)*, pages 139–148, New York, 1989.
- [73] D.Geiger and A.Shaffer. *Fourth Conference on Uncertainty in Artificial Intelligence (UAI'88)*. Number 48. 1998.
- [74] D.Geiger, J.Pearl, and T.S.Verma. Identifying independance in bayesian networks. *Networks*, 20 :507–534, 1988.
- [75] J.Pearl and T.S.Verma. Causal networks : Semantics and expressiveness. In *Fourth Workshop on Uncertainty in Artificial Intelligence*, pages 352–359. Mountain viex,CA, 1988.
- [76] J. Kim and J. Pearl. A computational model for causal and diagnostic reasoning in inference systems. In *Proceedings of Seventeenth International Joint Conference on Artificial Intelligence (IJCAI'83)*, pages 190–193, 1983.

- [77] J. Pearl. Fusion propagation and structuring in belief networks. In *Artificial Intelligence*, pages 241–288, 1986.
- [78] G.F.Cooper. Computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, 20 :661–685, 1990.
- [79] E.Castillo, J.M.Gutiérrez, and A.S.Hadi. *Expert systems and probabilistic network models*. Springer, 1997.
- [80] M.I.Jordan, K.P.Murphy, and Y.Weiss. Loopy belief propagation for approximate inference : An empirical study. In *Fifteenth Conference on Uncertainty in Artificial Intelligence UAI99*, pages 467–475, 1999.
- [81] M.A.Peot and R.D.Shachter. Fusion and propagation with multiple observations in belief networks. *Artificial Intelligence*, 48 :299–318, 1991.
- [82] A.P.Dawid, S.L.Lauritzen, B.N.Larsen, and H.G.Leimer. Independence proprieties of directed markov fields. *Networks*, 20 :491–505, 1990.
- [83] M.Yannakakis. Computing the minimum fill-in is np-complete. *SIAM Journal on Algebrics and Discrete Methods*, 2 :77–79, 1981.
- [84] S.Arnborg, D.G.Corneil, and A.Prokurowski. Complexity of finding embeddings in a k-tree. *SIAM Journal on Algebrics and Discrete Methods*, 8 :277–284, 1987.
- [85] U.Kjærulff. Triangulation of graphs- algorithms giving small total state space. Technical report, Departement of Mathematics and computer Science, Aalborg University, Denmark, 1990.
- [86] C.Golumbic. *Triangulated graphs*, chapter Algorithms graph theory and perfect graphs, pages 98–100. Academic Press, New York University, 1980.
- [87] F.V.Jensen and F.Jensen. Optimal junction trees. In *Tenth Conference on Uncertainty in Artificial Intelligence*, pages 360–366. Seatle, Wash, 1994.
- [88] A.Darwiche and C.Huang. Inference in belief networks : A procedural guide. *International Journal of Approximate Reasoning*, 15 :225–263, 1996.
- [89] F.V.Jensen, S.L.Lauritzen, and K.G.Olesen. Bayesian updating in causal probabilistic networks by local computations. *Comput.Statist.Quart*, 4 :269–292, 1990.
- [90] C.Borgelt and R.Kruse. Some experimental results on learning probabilistic and possibilistic networks with different evaluation measures. In *European Conference of Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECS-QARU97)*, pages 71–85. Bad Honnef (Germany), 1997.
- [91] S.A.Cook. The complexity of theory -proving procedures. In *Third Annual ACM Symposium on Theory of Comuting*, pages 151–158, ACM New york, 1971.
- [92] K.F Khellaf-Haned. *Autour des modes de représentation et de raisonnement en théorie des possibilités quantitative*. PhD thesis, LRIA-Département d’Informatique-FEI-USTHB et CRIL-Université d’Artois, 2007.
- [93] C. Cayrol and M. Lagasque-Schiex. On the complexity of non-monotonic entailment in syntax-based approaches. In *ECAI94 Workshop on Algorithms, Complexity and Common-sens reasoning Amsterdam*, 1994.