

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITÉ ABOU BEKR BELKAID  
FACULTÉ DES SCIENCE  
DÉPARTEMENT INFORMATIQUE

## **THÈSE**

pour obtenir le grade de

**DOCTEUR**

Spécialité : **Informatique**

présenté et soutenu publiquement

par

**Youcef BENMOUNA**

Le 13 juin 2019

Titre:

# **Amélioration des performances des réseaux bayésiens dans le domaine médical**

## **Jury**

Président du Jury. Mr. Abdelkrim BENAMAR ,	MCA, UABB Tlemcen
Directeur de Thèse. Mr. Chikh Mohamed Amine,	Professeur, UABB Tlemcen
Co-Directeur de Thèse. Mr. Said MAHMOUDI,	Professeur, UMons Belgique
Examineur. Mr. Badr BENMAMMAR,	MCA, UABB Tlemcen
Examineur. Mr. Ghalem BELALEM,	Professeur, Université d'Oran
Examineur. Mme. Djamila HAMDADOU,	Professeur, Université d'Oran
Invité d'honneur. Mr. Mourtada BENZAOUZ ,	MCB, UABB Tlemcen



À mes parents,  
à ma femme ,  
à mes enfants : Ryad et Aya,  
à toute ma famille. . .

# Remerciements

Si on suppose que la thèse marque la fin des études, il serait affable d'être reconnaissant envers tous mes enseignants et témoigner de la gratitude à l'ensemble des personnes ayant participé à mon apprentissage depuis le premier instituteur qui m'a appris à lire ! Cependant et pour des raisons pratiques, je me limiterai à ne citer que les personnes qui ont contribué de façon directe au travail présenté dans ce manuscrit. Aussi bien au niveau scientifique qu'au niveau humain.

Tout d'abord, Je tiens à remercier en premier lieu Mr. Mohamed Amine CHIKH, mon directeur de thèse pour sa sympathie, sa disponibilité, ces idées, ces conseils et ces encouragements qui m'ont permis de mener à bien cette thèse.

J'adresse aussi mes très sincères remerciements à Mr. Abdelkrim BENAMMAR de me faire l'honneur de s'intéresser à ce travail et d'avoir présidé le jury.

J'exprime ma profonde reconnaissance à Mr. Badr BENMAMMAR, Mr. Ghalem BELALEM Professeur, Université d'Oran, Mme. Djamila HAMDADOU Professeur, Université d'Oran, Mr. Mourtada BENAZZOUZ et Mr. Said MAHMOUDI, Professeur, UMons Belgique pour l'intérêt qu'ils ont bien voulu porter à ce travail en acceptant de l'examiner et d'en être rapporteurs.

L'ensemble des travaux présentés dans ce mémoire n'auraient pas vu le jour sans le soutien et la participation du Mr Fethallah HADJILA, par sa contribution à la rédaction de mes papiers scientifiques. Et pour finir un grand merci à mes collègues enseignants au département d'informatique.

# Table des matières

Remerciements . . . . .	i
<b>Remerciements</b>	<b>i</b>
Table des matières . . . . .	ii
Table des figures . . . . .	v
Liste des tableaux . . . . .	vii
Notations . . . . .	ix
<b>Notations</b>	<b>ix</b>
<b>Introduction</b>	<b>1</b>
<b>1 Réseau Bayésien</b>	<b>5</b>
1 Introduction . . . . .	5
2 Incertitude . . . . .	6
2.1 Définition . . . . .	6
2.2 Typologie de facteurs d'incertitude . . . . .	7
2.2.1 Générateurs d'incertitude liés au sujet . . . . .	7
2.2.2 Générateurs d'incertitude liés à l'objet . . . . .	9
2.2.3 Générateurs d'incertitude liés au contexte . . . . .	9
3 Décision médicale . . . . .	10
3.1 Différentes formes d'incertitudes liées au processus des soins	10
3.2 Raisonnement clinique . . . . .	11
4 Réseaux Bayésiens . . . . .	12
4.1 Introduction . . . . .	12
4.2 Définition . . . . .	12
4.3 Exemple . . . . .	13
4.4 Propriétés des réseaux bayésiens . . . . .	14

4.4.1	Condition de markov . . . . .	14
4.4.2	D-séparation . . . . .	15
4.4.3	Cartes d'indépendance . . . . .	17
4.4.4	Equivalence de markov . . . . .	18
4.5	Construction d'un réseau bayésien . . . . .	18
4.5.1	Discrétisation . . . . .	19
4.5.2	Apprentissage de la structure . . . . .	20
4.5.3	Apprentissage des paramètres . . . . .	27
5	La Méthode Branch & Bound ( <i>B&amp;B</i> ) . . . . .	29
5.1	Introduction . . . . .	29
5.2	Algorithme général . . . . .	30
5.3	Problème de voyageur en commerce (PVC) . . . . .	31
5.4	Parallèle Branch&Bound . . . . .	35
5.4.1	Modèle parallèle multi-paramétrique . . . . .	35
5.4.2	Evaluation parallèle du modèle des bornes . . . . .	36
5.4.3	Modèle d'exploration d'arborescence parallèle . . . . .	37
6	Conclusion . . . . .	38
<b>2</b>	<b>Apprentissage de Structure</b>	<b>39</b>
1	Introduction . . . . .	39
2	Méthodes approximatives . . . . .	39
2.1	Heuristiques . . . . .	39
2.1.1	Algorithmes d'indépendance conditionnelle . . . . .	40
2.1.2	Algorithmes basés sur les scores . . . . .	42
2.2	Métaheuristiques . . . . .	43
3	Méthodes exactes . . . . .	46
3.1	Programmation dynamique . . . . .	47
3.2	Algorithme de recherche $A^*$ . . . . .	49
3.3	Programmation lineaire . . . . .	50
4	Conclusion : . . . . .	53
<b>3</b>	<b>Application des Réseaux Bayésiens dans le Domaine Médical</b>	<b>61</b>
1	Introduction . . . . .	61
2	Etat de l'art . . . . .	61
3	Conclusion . . . . .	72

<b>4</b>	<b>Approches Proposées</b>	<b>73</b>
1	Introduction . . . . .	73
2	Problématique . . . . .	73
3	Contributions . . . . .	79
3.1	Introduction . . . . .	79
3.2	Strategies du parcours et du branchement . . . . .	80
3.2.1	Exemple illustratif . . . . .	80
3.2.2	Algorithme Proposé . . . . .	86
3.2.3	Expérimentations . . . . .	88
3.2.4	Discussions . . . . .	92
3.3	Strategié de bornage . . . . .	96
3.3.1	Introduction . . . . .	96
3.3.2	Opérateur de bornage . . . . .	96
3.3.3	Algorithme proposé . . . . .	101
3.3.4	Expérimentations . . . . .	102
3.3.5	Discussions . . . . .	103
3.4	Parallel-CB-B&B Algorithm . . . . .	105
3.4.1	Introduction . . . . .	105
3.4.2	Algorithme parallèle . . . . .	105
3.4.3	Exprérimentations . . . . .	107
3.4.4	Discussions . . . . .	108
4	Conclusions . . . . .	109
	<b>Conclusion Générale</b>	<b>110</b>
	<b>Bibliographie</b>	<b>113</b>
<b>A</b>	<b>Bases de données</b>	<b>123</b>
1	Introduction . . . . .	123
2	Le cancer du sein . . . . .	124
2.1	Définition . . . . .	124
2.2	Caractéristique du noyau d'une cellule . . . . .	125
3	le Cardiotographie . . . . .	129
3.1	Déscription de la base de Données CTG . . . . .	131
4	Hépatite . . . . .	139

# Table des figures

1.1	Relations des incertitudes aux possibilités d'échecs . . . . .	6
1.2	Éléments clés de la définition proposée pour l'incertitude. [40] . . .	6
1.3	Cancer métastatique . . . . .	13
1.4	Exemple des connexions et séparations conditionnelles. . . . .	17
1.5	Exemple de cartes d'indépendances . . . . .	17
1.6	Phase de construction d'un réseau bayésien . . . . .	18
1.7	Cancer métastatique . . . . .	28
1.8	Exemple de PVC. . . . .	32
1.9	representation graphique de la méthode B&B pour la résolution du probleme de PVC. . . . .	34
1.10	Illustration du Modèle parallèle multi-paramétrique . . . . .	35
1.11	Illustration de l'Evaluation parallèle du modèle des bornes . . . . .	36
1.12	Illustration du modèle d'exploration d'arborescence parallèle. . . .	37
2.1	Exemple de recherche gloutonne. . . . .	56
2.2	Ensemble des sous-reseaux d'un réseau bayésien à quatre nœuds .	58
2.3	Graphe d'ordre pour quatre variables . . . . .	59
2.4	Parents possibles de la variable1 . . . . .	60
2.5	Graphe avec 3 nœuds . . . . .	60
4.1	Graphe initial G. . . . .	76
4.2	Execution de la méthode B&B. . . . .	82
4.3	Elimination des cycles selon la nouvelle startégie du parcours. . . .	83
4.4	Elimination des cycles selon la nouvelle startégie du branchement.	84
4.5	Elimination des cycles selon les deux startégies. . . . .	85
4.6	Structure de donnée Utilisée . . . . .	86

4.7	Nombre et Taille maximal de cycles que peut contenir un graphe durant de l'exécution de l'algorithme. . . . .	94
4.8	Exemple d'un ensemble des Cycles Disjoint par leurs Sommet . . .	97
4.9	Exemple deux Cycles avec un Sommet en commun . . . . .	98
4.10	Exemple deux Cycles avec un arc en commun . . . . .	98
4.11	Temps d'exécution de la méthode <i>Parallel CB-B&amp;B</i> , pour l'ASRB appliqué aux bases de données médicales. . . . .	107
4.12	Taux d'accélération de la méthode <i>Parallel CB-B&amp;B</i> , pour l'ASRB appliqué aux bases de données médicales. . . . .	108
A.1	Limites initiales approximatives des noyaux de cellules . . . . .	125
A.2	<i>snakes</i> après la convergence vers les noyaux cellulaires . . . . .	126
A.3	Lignes radiales utilisées pour le calcul du lissage . . . . .	127
A.4	Accords utilisés pour calculer la concavité . . . . .	128
A.5	Segments utilisés dans le calcul de symétrie . . . . .	128
A.6	Séquence de mesures pour calculer la dimension fractale . . . . .	129
A.7	Un enregistrement CTG typique comprenant FHR et UC. . . . .	130
A.8	Valeur de base de la fréquence cardiaque fœtale . . . . .	132
A.9	Les contractions utérines . . . . .	134
A.10	La décélération tardive . . . . .	135
A.11	La décélération variable : . . . . .	136
A.12	La décélération prolongée : . . . . .	137
A.13	La variation de la fréquence cardiaque fœtale à court terme : . . . .	138
A.14	La variation de la fréquence cardiaque fœtale à court terme : . . . .	139

# Liste des tableaux

1.1	Table de probabilité conditionnelle du nœud <b>Coma</b> . . . . .	14
1.2	Distribution de probabilité définie sur deux variables X et Y . . . . .	17
1.3	Base d'exemple du nœud <b>H</b> . . . . .	29
1.4	TPC pour la variable <b>H</b> . . . . .	29
1.5	Matrice d'adjacence représentant les poids. . . . .	32
2.1	Codage des variables de famille d'un graphe avec 3 nœuds . . . . .	60
4.1	Cache des Scores locaux pour chaque variable . . . . .	75
4.2	Matrice A . . . . .	76
4.3	Comparaison entre la méthode Standard de B&B et la méthode B&B améliorée . . . . .	86
4.4	Bases de données utilisées dans les expérimentations . . . . .	89
4.5	Taille du Cache et de l'Espace de Recherche . . . . .	89
4.6	Résultats (taille de la mémoire & nombre d'itérations) pour obtenir la première solution par le nouveau B&B (respectivement le stan- dard B&B). . . . .	91
4.7	Ratios comparatives pour obtenir la première solution . . . . .	91
4.8	Ratios comparatives pour obtenir la solution optimale . . . . .	92
4.9	Résultats (taille de la mémoire & nombre d'itérations) pour obtenir la solution optimale par le nouveau B&B (respectivement le stan- dard B&B). . . . .	92
4.10	Temps d'exécution du nouveau <i>B&amp;B</i> vs celui de différents algo- rithmes de l'état de l'art. Les temps sont donnés en secondes . . . . .	93
4.11	Comparaison entre <i>CB – B&amp;B</i> et B&B en termes de mémoire . . . . .	103
4.12	Comparaison entre <i>CB – B&amp;B</i> et B&B en termes de nombre d'ité- rations . . . . .	103

4.13 Temps d'exécution de CB-B&B par rapport aux différents algorithmes de l'état de l'art. Le temps d'exécution est en millisecondes ( <i>Fail</i> signifie que la méthode n'a pas pu résoudre le problème dans un délai de deux heures) . . . . .	104
---	-----

# Notations

IA : Intelligence Artificielle

ASRB : Apprentissage de Structure des Réseaux Bayésiens

GOSC : Graphe Orienté Sans Circuits

TPC : Tables de Probabilités Conditionnelles

AIC : Critère d'Information d'Akaike

BIC : Critère d'Information Bayésien

BD : Score Bayésien

BDe : Score Bayésien de Dirichlet équivalent

BDeu : Score Bayésien de Dirichlet équivalent et uniforme

fNML : Maximum de Vraisemblance Normalisé et factorisé

MDL : Minimum description length

B&B : Branch&Bound

POC : Problème d'Optimisation Combinatoire

PVC : problème de voyageur en commerce

ARPM : Arbre de Recouvrement de Poids Minimal

PC : Algorithme de Peter et Clark

IC : Inductive Causation

RG : Recherche Gloutonne

MTC : Médecine Traditionnelle Chinoise

GA : Algorithme Génétique

ACO : méthode de la colonie de fourmis

PSO : Algorithmes de particules d'essaim d'optimisation

CPSO : Algorithmes chaotiques de particules d'essaim d'optimisation

VO : Ordre des Variables

VOEA : Algorithme d'Evolution à Classement des Variables

DMO : Opérateur de Mutation basé sur la Distance

FEP : Fonction d'Étalement Ponctuel

RBD : Réseaux Bayésiens Dynamiques

NB : Naïve Bayes

BNF : Bayésien Naïf Flou

DORB : Diagnostic médical basé sur l'Ontologie et le Réseau Bayésien

WDBC : Wisconsin Diagnostic Breast Cancer

HARM : Hierarchical Association Rule Model

CoD : Causes de Décès

CTG : Cardiography

# Introduction

Dans le domaine médical, une décision fait face à un processus complexe, impliquant différents mécanismes de pensée clinique et de prise de décision, ces mécanismes sont affectés par des biais potentiels ou des exogènes. Le diagnostic médical peut être considéré comme une activité intellectuelle, où le médecin synthétise les informations obtenues dans la situation clinique, les intègre aux connaissances et aux expériences antérieures et les utilise dans le but de réaliser une prise de décision. Ces processus de réflexion et de prise de décision sont au cœur de la pratique professionnelle.

Le challenge principal de la prise de décision médicale, est dans la nature de l'incertitude à laquelle le décideur est confronté et de la manière dont celui-ci l'interprète. Cette incertitude provient d'une part, du malade duquel le risque est associé, et du médecin traitant, d'autre part. La pratique du médecin se trouve face à des situations où il ne dispose pas toutes les informations nécessaires permettant de poser un diagnostic et prendre la meilleure décision. Donc il est quotidiennement confronté à l'incertitude. Chaque situation possède un ensemble d'issues possibles, choisir la meilleure représente un défi pour le médecin, où il doit attribuer à chaque issue un degré de vraisemblances et d'estimer l'ampleur des conséquences.

L'intelligence artificielle(IA) consiste à utiliser des techniques permettant de simplifier la tâche du médecin, voire de "remplacer" l'intervention humaine dans le raisonnement clinique. L'IA est la partie informatique dédiée à la conception des systèmes informatiques intelligents, capables de tirer des conclusions pertinentes sans intervention directe du médecin. Ces systèmes sont basés sur des modèles construits pour chaque pathologie, issus des rapports entre ces données

et les signes cliniques que présentent les patients.

Parmi les techniques d'IA utilisées dans telle situation, les Réseaux Bayésiens présentent une alternative très intéressante, ils apportent une meilleure modélisation à l'aspect incertain qui relie les différentes variables du domaine médical ; ils représentent les relations entre ces variables d'une façon visuelle, en plus, ce modèle est enrichi par des distributions de probabilités permettant de quantifier le taux de certitude de ces relations.

La tâche la plus difficile dans la construction d'un réseau bayésien, est l'apprentissage de structure, à cause de la complexité de l'espace de recherche, sachant qu'un nombre limité de variables peut avoir un nombre exponentiel de différents modèles possibles. Dans la littérature, ils existent plusieurs méthodes pour l'Apprentissage de Structure des Réseaux Bayésiens(ASRB), réparties en deux grandes classes : *approchées et exactes*, Les méthodes approchées produisent une solution réalisable dans un temps raisonnable, et les autres cherchent la solution optimale et peuvent prendre un temps exponentiel.

### **Contributions**

Dans un but d'optimiser le raisonnement clinique, nous nous intéressons dans ce travail à une approche exacte qui est la méthode standard Branch&Bound [37] pour l'ASRB, où nous l'avons analysé et amélioré en proposant des nouvelles stratégies. Cette méthode est appliquée à un graphe cyclique, où à chaque itération un cycle est éliminé jusqu'à l'obtention des graphes acycliques, le meilleur modèle parmi ces graphes est considéré comme solution au problème donné.

L'objectif de cette thèse est d'améliorer l'apprentissage de structure des réseaux bayésiens. Où, nous décrivons en détails l'algorithme standard *Branch&Bound* appliqué à l'ASRB, ainsi, que ces avantages par rapport à d'autres méthodes exacts. Puis, nous proposons de nouvelles heuristiques, permettant d'améliorer considérablement ce dernier.

Nous proposons quatre contributions, permettant d'améliorer les performances

de cette méthode :

1. La méthode standard utilise *le meilleur d'abord* comme stratégie du parcours, cette stratégie produit moins de solutions intermédiaires. Nous adoptons une combinaison de deux conduites ; la recherche en *profondeur* et du *meilleur d'abord*.
2. La procédure du branchement appliqué par cette méthode à un graphe, ne prend pas en considération la complexité spatiale. Nous proposons alors une nouvelle procédure qui permet d'optimiser au maximum la complexité spatiale.
3. L'algorithme standard B&B utilise le score  $S$  du graphe  $G$  comme borne supérieure. Pour accélérer l'exploration d'arbres de recherche, nous utilisons un opérateur efficace pour le calcul de cette borne.
4. Le parallélisme est adopté pour l'implémentation de notre nouvel algorithme.

Pour valider nos contributions, nous avons réalisé plusieurs expérimentations qui nous ont illustrés une nette progression des performances, en les comparants avec les résultats obtenus avec l'algorithme standard, ainsi qu'avec d'autres méthodes exacts existantes dans la littérature.

### **Organisation de la thèse :**

Cette thèse s'articule autour de quatre chapitres :

Le premier chapitre présente le concept de l'incertitude, ses typologies ainsi que son implication dans différents domaines. Ce chapitre aborde également la décision médicale qui se base sur des données prises des sources différentes, qui sont généralement incertaines. Par la suite nous décrivons les réseaux Bayésiens, leurs propriétés et les étapes de construction de ces modèles. Á la fin de ce chapitre nous présentons la méthode Branch&Bound que nous avons adopté pour la construction de structure du Réseau Bayésien.

Le deuxième chapitre est consacré pour les différents algorithmes utilisés pour

l'apprentissage de structure des réseaux bayésiens. Où, nous présentons quelques méthodes approximatives et d'autre exactes avec plus de détails.

Le troisième chapitre décrit quelques travaux proposés dans le domaine médical, où nous nous sommes intéressés à l'implication des réseaux bayésiens dans le processus décisionnel. Ces techniques ont pour objectif la construction des modèles à partir des données requises et des experts, pour aider le médecin à prendre des meilleures décisions.

Le quatrième et dernier chapitre est destiné à la présentation de nos contributions ; nous commençons par une description détaillée de l'algorithme *B&B* standard, dédié pour l'ASRB, ensuite, nous montrons ses limitations afin de proposer de nouvelles améliorations. Puis nous exposons la première et la deuxième contribution consacrées aux stratégies du parcours et de branchement. Juste après l'opérateur du bornage et le procédé du parallélisme constituent le fondement de la troisième et quatrième contribution. Chacune de ces propositions est sanctionnée par des expérimentations et des discussions.

Finalement, une conclusion termine ce document et rappelle les principales contributions élaborées tout au long de ce travail. Elle présente également les perspectives relatives aux résultats obtenus.

# Chapitre 1

## Réseau Bayésien

### 1 Introduction

Dans le domaine médical, le praticien est confronté quotidiennement à des situations d'incertitude. Il se trouve face à des cas où il ne dispose pas toutes les informations nécessaires permettant de prendre la meilleure décision.

Chaque situation a un ensemble d'issues possibles, choisir la meilleure représente un challenge pour le médecin.

Parmi les techniques d'intelligence artificielle utilisées pour aider les médecins à prendre des meilleures décisions dans telles situations, les réseaux Bayésiens constituent un choix très intéressant, car l'incertitude est prise en compte dans son raisonnement, en plus d'un graphe causal permettant de représenter la connaissance d'une façon lisible.

Dans ce chapitre nous allons décrire l'incertitude et ses typologies, ainsi que la décision médicale qui se base sur des données prises des sources différentes, généralement incertaines. Par la suite nous présentons les réseaux Bayésiens, ses propriétés et les étapes de construction de ce modèle. A la fin de ce chapitre nous décrivons la méthode Branch&Bound (B&B) que nous avons adopté pour la construction de structure de Réseau Bayésien.

## 2 Incertitude

Dans la vie quotidienne, nous sommes souvent confrontés à des situations où l'incertitude règne. Malgré son implication dans différents domaines, il n'existe pas une définition commune de ce concept.

### 2.1 Définition

FIGURE 1.1 – Relations des incertitudes aux possibilités d'échecs

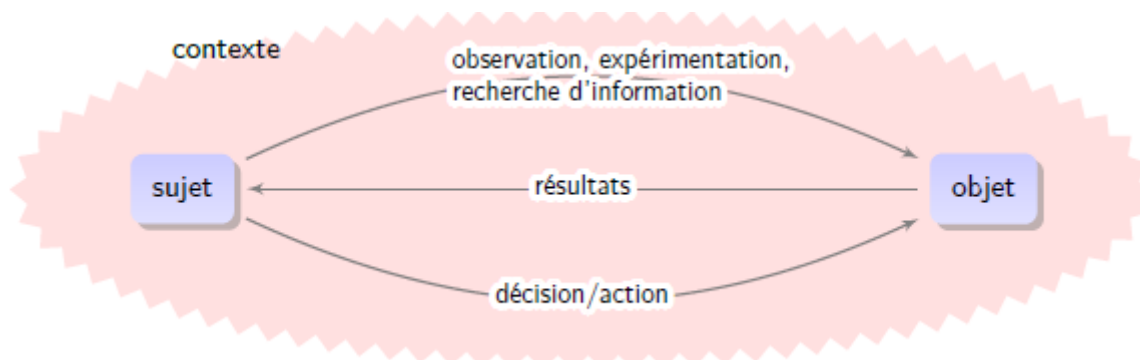


D'après [40] « L'incertitude est un manque conscient de connaissance d'un sujet, relative à un objet, non encore parfaitement défini, dans un contexte nécessitant une décision/action. »

Le degré d'incertitude est fortement lié aux possibilités d'échecs ou parfois tout devient ambigu [41]. (figure 1.1).

La figure 1.2 est une illustration de cette définition. Où, les flèches représentent les tentatives d'acquisition de connaissances, de la part du sujet, avec des

FIGURE 1.2 – Éléments clés de la définition proposée pour l'incertitude. [40]



observations, d'expérimentations, de recherche d'information. Après cette tentative, les résultats provenant de l'objet, sont récupérés et interprétés par le sujet. Finalement, ce dernier prend un ou plusieurs actions/décisions du sujet sur la base de ces résultats. Ce processus peut être répété jusqu'à ce que le sujet soit satisfait de son état de connaissances ou de ses actions/décisions.

La prise de décision devient facile, dans le cas où le sujet connaît exactement l'état présent et futur de l'objet, et toutes les options d'action que l'objet lui donne, donc il ne sera pas confronté à l'incertitude, c.à.d. le décideur possède toutes les informations nécessaires sur les gains et les pertes relatif à chaque décision.

l'incertitude est caractérisé par :

- le manque de connaissance ;
- le contexte nécessitant une action/décision.

L'incertitude signifie que les données disponibles ne permettent pas d'expliquer une situation aussi précisément qu'on le voudrait ; ce n'est pas une ignorance, c'est une imprécision. Il est donc important de clarifier le plus possible les limites des données existantes, afin d'en faire la meilleure utilisation [95].

## 2.2 Typologie de facteurs d'incertitude

Pour mieux cerner les contours d'une incertitude, il faudrait pouvoir mieux identifier ses propres sources, qui peuvent provenir du sujet, de l'objet ou du contexte [40].

### 2.2.1 Générateurs d'incertitude liés au sujet

Les générateurs liés au sujet peuvent être divisés en deux catégories : individuels et collectifs. Chaque sujet réagit d'une manière différente. Des propriétés de la personnalité du sujet peuvent être une source d'incertitude. En plus de ses expériences et ses qualités qui jouent aussi un rôle important dans la génération individuel de l'incertitude.

Le sujet est en interaction continue avec les autres éléments du groupe. Ces

échanges influent sur le comportement du sujet en lui offrant plus ou moins de connaissance ou de doute.

**a-Générateurs individuels** : Les générateurs d'incertitude individuels sont soit des propriétés psychologiques du sujet, soit un manque de compétences et d'expériences qui génère de l'incertitude :

- **doute** : état de l'esprit qui questionne, sous forme de remise en cause. Un médecin peut avoir un doute lors de la prescription médicale, à cause d'un manque d'informations issues des résultats d'analyses.
- **hésitation** : manque d'assurance. Le sujet n'est pas convaincu de sa décision, ce qui lui rend incapable de la communiquer.
- **scepticisme** : attitude, disposition d'esprit d'une personne qui ne croit pas ou ne fait pas confiance à ses opinions et à ses valeurs. C'est une personnalité sceptique par nature, qui remettent en question les résultats peu concluants.
- **irrésolution** : personnalité individuelle qui n'est pas valable pour une décision, quel que soit le contexte.
- **Fréquence** : État mental d'une personne souffrant de difficultés d'autodétermination.
- **Pessimisme** : comportement rationnel de ne voir que le mauvais côté des choses, de constater que tout va mal. Personne n'a tendance à exagérer les résultats négatifs et leurs conséquences sur une telle décision.
- **Aversion au risque** : c'est une personnalité qui à une peur excessive de risque, ce qui entraîne une méfiance excessive et disproportionnée qui entraîne la paralysie de la décision. Dans ce cas, la prime de risque sera la différence entre les objectifs attendus et les résultats réels au cours de risque potentiel. Par conséquent, le risque est considéré comme une menace. La personne participe en permanence à un principe de précaution dans les décisions.
- **Fonction de regret** : La tendance à remettre en cause la décision adoptée, préférant l'option non retenue dans la décision.

**b-Générateurs collectifs** : L'interactions entre les individus qui ont une influence sur la prise de décision de l'individu peuvent augmenter l'incertitude :

- **débats contradictoires** : ces types de débats entre les décideurs produisent des résultats divergents, qui influent sur l'incertitude.

- **influences d'expertises** : l'avis d'un expert peut influencer par sa compétence, qui peut avoir une interprétation des résultats qui oppose les avis des décideurs internes.
- **relation subordonné entre individus** : La position hiérarchique d'un individu peut changer l'avis de ses collègues.
- **divergence culturel** : Dans le cas ou dans un comité de décision, deux groupes d'individus s'opposent, cette situation cause une perturbation de l'évaluation des risque, car, les personnes, appartenant au groupe X, veulent être solidaires entre elles, au lieu de regarder, objectivement au interprétations.
- **effets de groupe** : l'individu adopte le résultat d'un projet, proposé par la majorité du groupe, sans être convaincu par la pertinence de cette évaluation.

### 2.2.2 Générateurs d'incertitude liés à l'objet

Ce sont des propriétés de l'information caractérisant l'objet, qui ne sont pas suffisantes pour comprendre le comportement de cet objet :

- **absence** : le manque d'information.
- **incomplétude** : l'incomplétude des informations sur certaines caractéristiques du système, qui peut être dues à l'impossibilité d'obtention de certains renseignements, au moment de prélèvement. Elle peut aussi dues à l'existence d'informations générales sur l'état de tout le système.
- **ambiguïté** : cette ambiguïté est causé par des interprétations multiples et conflictuelles, entraînant un trouble et une insuffisance de compréhension.
- **contradiction** : l'existence des informations contradictoires sur le même objet.
- **multidisciplinarité** : ce sont des informations qui touchent plusieurs domaines à la fois, provoquant des difficultés de compréhension.
- **volatilité** : l'instabilité d'une valeur dans le temps.

### 2.2.3 Générateurs d'incertitude liés au contexte

Ces générateurs sont des variables qui concernent l'environnement, la méconnaissance de ces variables génère de l'incertitude, ils ont aussi un impact sur la décision.

### 3 Décision médicale

Le défi principal du problème de décision en général et dans le domaine médical en particulier, est dans la nature de l'incertitude à laquelle le décideur est confronté et de la manière dont celui-ci l'interprète [96].

La pratique du médecin se trouve face à des situations où il ne dispose pas toutes les informations nécessaires permettant de poser un diagnostic et prendre la meilleure décision. Chaque médecin est donc confronté quotidiennement à ces situations d'incertitude. Chaque situation a un ensemble d'issues possibles, choisir la meilleure issue représente un challenge pour le médecin, où il doit attribuer à chaque situation un degré de vraisemblances et d'estimer l'ampleur des conséquences.

Dans le domaine de l'incertitude médicale, il existe deux typologies. D'une part en fonction de la personne à laquelle le risque est associé : le patient (la demande) comme le médecin (l'offre) subissent un aléa, d'autre part en fonction de l'étape du processus de "soins" : étape de prévention de la maladie (prévention primaire ou passive), étape de dépistage de la maladie (prévention secondaire ou active), étape de diagnostic, et enfin étape de soins curatifs.

#### 3.1 Différentes formes d'incertitudes liées au processus des soins

On distingue trois formes d'incertitudes liées au processus des soins [42] :

- **La première est celle de la qualité du diagnostic** : Le rôle du diagnostic médical consiste à collecter toute information nécessaire au moyen des questions ou bien des analyses cliniques. Néanmoins, l'information collectée au cours de ce processus est généralement incomplète puisque les enquêtes faites par le médecin ne sont pas exhaustives, et orientées par les patient, alors que, le patient n'est pas un expert dans le domaine médical, il peut donc négliger, volontairement ou non, quelques caractéristiques de son état, comme, il peut influencer le diagnostic du médecin en attirant son attention

sur une information complètement erronée. Pour répondre à ce type d'incertitude, le médecin doit multiplier les examens complémentaires et les analyses cliniques au détriment de l'écoute de patient.

- **La seconde incertitude porte sur le traitement prescrit :** Ce traitement peut être inefficace, même si le diagnostic est correct. Le médecin n'assure donc que le traitement curatif qu'il prescrira, sera bien adapté à son patient . Dans le cas où les soins curatifs sont efficaces, l'incertitude se déplace sur la qualité ces soins. C.à.d., pour une pathologie donnée, les prescriptions proposés par le médecin sont multiples, le patient espère non seulement d'être guéri, mais de l'être dans les meilleures situations possibles. Or les critères de mesure de qualité sont aussi de nature probabiliste. Ici encore, il n'y a pas une garantie que le traitement choisi par le médecin soit le meilleur.
- **La troisième incertitude concerne la connaissance même d'une thérapie ou d'une maladie :** En matière médicale, les états de la nature ne sont jamais complètement connus. La liste des pathologies, comme celle des thérapies, est momentanée. Le praticien est donc structurellement incapable d'évaluer le risque d'une telle maladie, même si le diagnostic est exhaustif. Supposons le cas contraire, le praticien ne puisse jamais juger d'un éventuel risque futur.

Deux aspects caractérisent l'incertitude médicale. Qui sont principalement essentielles en raison d'une part d'une incomplétude des états de la nature (maladie, traitement, mais aussi diagnostic), d'autre part en raison de son facteur de risque.

### 3.2 Raisonnement clinique

Le raisonnement clinique est le processus décisionnel qui permet au médecin de prendre les actions les plus appropriées dans un contexte spécifique pour résoudre les problèmes de santé. On peut le considérer comme une activité intellectuelle ; le médecin synthétise les informations obtenues dans la situation clinique, les intègre aux connaissances et à l'expérience antérieures et les utilise dans le but dans une utilisation pour une prise de décision [43,45,45].

Ces processus de réflexion et de prise de décision sont au cœur de la pratique

professionnelle.

L'intelligence artificielle(IA) consiste à utiliser des techniques pour remplacer la cognition humaine dans le raisonnement clinique. En d'autres termes, l'IA est la partie informatique dédiée à la conception de systèmes informatiques intelligents, capables de tirer des conclusions pertinentes sans intervention directe du médecin. Le raisonnement clinique se base généralement sur des données prise des sources différentes, qui sont généralement incertaines.

Parmi les techniques d'IA utilisées dans telle situation, les Réseaux Bayésiens constituent une alternative très intéressante, ils apportent une meilleure modélisation à l'aspect incertain qui relie les différents variables du domaine médical ; ils représentent les relations entre les variables du modèle d'une façon visuelle, ce modèle est enrichit par des distributions de probabilités permettant de quantifier le taux de certitude de ces relations.

## 4 Réseaux Bayésiens

### 4.1 Introduction

Les réseaux bayésiens reposent sur un formalisme basé sur les théories des probabilités et des graphes, ils représentent d'une façon visuelle les relations (dépendances et indépendances) entre les variables du modèle, ces relations sont indiquées par des arcs entre les sommets. De plus, les distributions de probabilités permettent d'enrichir cette structure graphique par une quantification de ces relations.

### 4.2 Définition

Un réseau bayésien [1] est un modèle graphique probabiliste, représenté par un ensemble de relations conditionnelles dans un domaine  $V = \{X_1, X_2, \dots, X_n\}$  de variables stochastique ayant chacune leur propre domaine de définition.

Il est défini par un couple  $(G, \theta)$  :

- $G$  est un Graphe Orienté Sans Circuits(GOSC), dont les sommets correspondent aux variables  $(X_1, X_2, \dots, X_n)$  de l'ensemble  $V$  et les arcs orientés de  $G$  représentent les dépendances entre ces variables.

- $\theta$  est l'ensemble des paramètres du réseau. Il contient les paramètres  $\theta_{i,j,k} = P(X_i = x_{ki} | PA_i = pa_{ij})$ ,  $i \in \{1 \dots n\}$  pour chaque valeur  $x_{ki}$  pouvant être prise par  $X_i$  et chaque configuration  $n_{ij}$  de  $PA_i$ , (ensemble des sommets parents de  $X_i$  dans  $G$ ).

### 4.3 Exemple

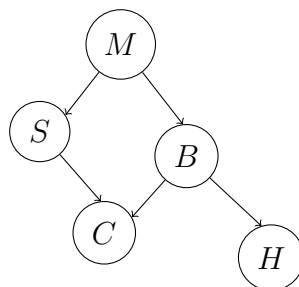


FIGURE 1.3 – Cancer métastatique

Dans cet exemple 1.3, le réseau Bayésien est utilisé dans le diagnostic médical (Cancer métastatique), ce modèle est représenté par un GOSC, où toutes les variables (nœuds) sont Booléennes.

Les relations conditionnelles sont représentées par des arcs entre les différents nœuds :

- Le cancer métastatique (M) est une cause possible de tumeurs cérébrales (B) et est également une explication pour l'augmentation du calcium sérique total (S).
- L'augmentation du calcium sérique total (S) pourrait expliquer un patient qui tombe dans un coma (C).
- Les tumeurs cérébrales (B) provoquent un grave mal de tête (H) et peuvent expliquer qu'un patient soit au coma (C).

Les paramètres de ce modèle sont définis par des Tables de Probabilités Conditionnelles (TPC) pour chaque sommet du graphe, qui caractérisent les dépendances stochastiques entre les variables du modèle.

Le graphe et les TPCs d'un RB peuvent être construits manuellement par un expert ou automatiquement extraites à partir des bases de données d'exemples

TABLE 1.1 – Table de probabilité conditionnelle du nœud **Coma**

S	B	P(C=Vrai S,B)	P(C=Faux S,B)
Vrai	Vrai	0.90	0.10
Vrai	Faux	0.80	0.20
Faux	Vrai	0.80	0.20
Faux	Faux	0.05	0.95

par des algorithmes d'apprentissage automatique.

Le modèle bayésien est utilisée pour déterminer les probabilités correspondantes à certains événements, certaines questions, certaines dépendances. Les inférences permettent effectivement de calculer la probabilité des données non observées. Par exemple, en fonction des symptômes d'un malade, on calcule les probabilités des différentes pathologies compatibles avec ces symptômes.

## 4.4 Propriétés des réseaux bayésiens

L'application des réseaux bayésiens [2] dans un domaine permet de lier les propriétés graphiques de la structure  $G$  d'un réseau bayésien  $B$  avec les propriétés de la distribution de probabilités modélisée.

Dans ce qui suit nous allons citer quelques propriétés :

### 4.4.1 Condition de markov

Un réseau bayésien  $(G, \theta)$  vérifie la condition de markov si chaque sommets  $X \in G$ , doit être indépendante de ses non descendantes dans  $G$  conditionnellement à ses parents [4,5].

La condition de Markov implique la factorisation de la loi jointe :

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i | PA_i^G). \quad (1.1)$$

La condition locale de Markov implique que deux sommets non adjacents  $X_i$  et

$X_j$  de  $G$  sont conditionnellement indépendants étant donné un troisième sous-ensemble, contenu dans  $U/\{X_i, X_j\}$ .

A partir de cette propriété, la table de probabilités doit contenir que les valeurs de  $P(X_i = x_i | Pa_i^G = pa_i)$  pour toutes les valeurs de  $x_i$  et les possibles instanciations conjointes de  $pa_i$ .

#### 4.4.2 D-séparation

L'interprétation d'un réseau bayésien se base sur un principe d'indépendance conditionnelle. La lecture de ces indépendances est liée à la notion de séparation [6].

Dans un GOSC, La d-séparation est un critère qui s'applique sur trois sous-ensembles de variables (le d de d-séparation provenant de l'orientation des arcs dans le graphe). Il détermine les indépendances conditionnelles modélisées par un GOSC.

Ce critère permet de vérifier si un sous-ensemble  $X$  de variables du domaine est conditionnellement indépendant d'un sous-ensemble  $Y$  sachant un sous-ensemble  $Z$ .

**Définition 1.1. (Connexion série)** Dans un graphe orienté  $G$ , une connexion est dite série, tout triplet de sommets  $\{A, B, C\}$  tel que :  $A \rightarrow B \rightarrow C$ .

**Définition 1.2. (Connexion divergente)** Dans un graphe orienté  $G$ , une connexion est dite divergente, tout triplet de sommets  $\{A, B, C\}$  tel que :  $A \leftarrow B \rightarrow C$ .

Dans ces types de connexions(série et divergente) :

- A et C sont dépendants ;
- A et C sont indépendants conditionnellement à B (si B est connu, A n'apporte aucune information à C) ;
- $P(C | A, B) = P(C | B) = P(C | \text{Parent}(C))$

**Définition 1.3. (V-structure)** Dans un graphe orienté  $G$ , une connexion est dite convergence (ou V-structure), tout triplet de sommets  $\{A, B, C\}$  tel que :  $A \rightarrow B \leftarrow C$ .

Dans ce cas :

- A et C sont indépendants ;
- A et C sont dépendants conditionnellement à B (si B est connu, A apporte une information à C) ;
- $P(C | A, B) = P(C | B) = P(C | \text{Parent}(C))$

**Définition 1.4. (Chemin)** Dans un graphe  $G$ , un chemin entre deux sommets  $A$  et  $B$  de  $G$  désigne une série d'arcs consécutifs reliant  $A$  à  $B$ , quelle que soit l'orientation des arcs.

**Définition 1.5. (Chemins bloqués)** Soit  $G = \{V, E\}$ , un graphe orienté sans circuit. Soit  $A, B$  et  $C$  trois sous-ensembles disjoints de  $V$ . Un chemin reliant un sommet de  $A$  à un sommet de  $B$  est dit bloqué par l'ensemble  $C$  si au moins une des deux conditions suivantes est vérifiée :

- Le chemin converge en un sommet  $X_i$  et ni  $X_i$  ni aucun de ses descendants ne sont dans  $C$  ;
- Le chemin passe par un sommet  $X_i$ , et  $X_i \in C$  en lequel il n'y a pas de convergence.

le chemin est dit actif si aucune de ces conditions n'est vérifiée.

**Définition 1.6. (d-séparation)** Soit  $G = \{V, E\}$ , un GOSC. Soit  $A, B$  et  $C$  trois sous-ensembles disjoints de  $V$ . On dit que  $A$  et  $B$  sont d-séparés par  $C$  dans  $G$  (noté  $(A \perp_G B | C)$ ) ssi tous les chemins qui relient un sommet de  $A$  à un sommet de  $B$  sont bloqués par  $C$ .

Si  $A$  et  $B$  ne sont pas d-séparés par  $C$ , on dit qu'ils sont d-connectés conditionnellement à un sous-ensemble  $C$ .

Considérant la figure 1.4, où les sommets grisés  $X_6$  et  $X_8$  appartiennent au sous-ensemble  $C$ , on remarque que :

- $X_3$  et  $X_5$  sont d-séparés (convergence en  $X_4$ ) ;
- $X_2$  et  $X_4$  sont d-connectés ;
- $X_7$  et  $X_9$  sont d-séparés par  $C$  ( $X_6$  appartient à  $C$ ) ;
- $X_1$  et  $X_3$  sont d-connectés par  $C$  ( $X_8$  appartient à  $C$ ).

FIGURE 1.4 – Exemple des connexions et séparations conditionnelles.

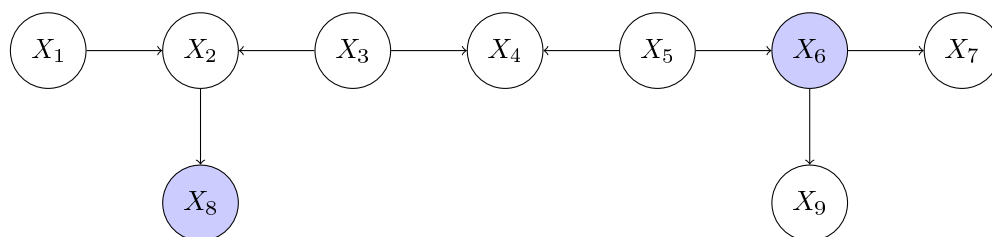


FIGURE 1.5 – Exemple de cartes d'indépendances



#### 4.4.3 Cartes d'indépendance

La qualification d'un graphe  $G$  par rapport à la distribution de probabilités  $P$  est vérifiée sur la base des relations de séparation détectées dans ce graphe.

Soit  $P$  une distribution de probabilités sur un ensemble de variables aléatoires  $V$ ;  $G$  un GOSC composé sur  $V$ ,  $A, B, C \subseteq V$ , et  $(\perp_G \cdot | \cdot)$  une relation d'indépendance conditionnelle vérifiée par  $P$ .

- $G$  est une carte d'indépendance (I-map) de  $P$  ssi :  $A \perp_G B | C \Rightarrow A \perp_P B | C$
- $G$  est une carte de dépendance (D-map) de  $P$  ssi :  $A \perp_P B | C \Rightarrow A \perp_G B | C$
- $G$  est une carte parfaite (P-map) de  $P$  ssi :  $A \perp_G B | C \Leftrightarrow A \perp_P B | C$

Soit l'exemple suivant, qui définit une distribution de probabilité  $P$  sur deux variables  $X$  et  $Y$ , on remarque que  $X \perp_P Y$  puisque :

$$\forall x, y \in \{Vrai, Faux\}^2 : P(X = x, Y = y) = P(X = x) \times P(Y = y)$$

TABLE 1.2 – Distribution de probabilité définie sur deux variables  $X$  et  $Y$ 

X	Y	P(X,Y)
Vrai	Vrai	0.48
Vrai	Faux	0.12
Faux	Vrai	0.32
Faux	Faux	0.08

Soit les deux graphes  $G1$  et  $G2$  illustrés dans la figure 1.5, on remarque que le

premier graphe  $G_1$  est une carte parfaite (P-map) de  $P$ . alors que  $G_2$  est un I-map de  $P$  (Tableau 1.2).

D'après la définition 1.3, un réseau bayésien est toujours une I-map de la distribution encodée. L'inconvénient majeur de cette définition est qu'un graphe qui n'a aucune indépendance (entièrement connecté) est aussi une I-map de la distribution qu'il représente.

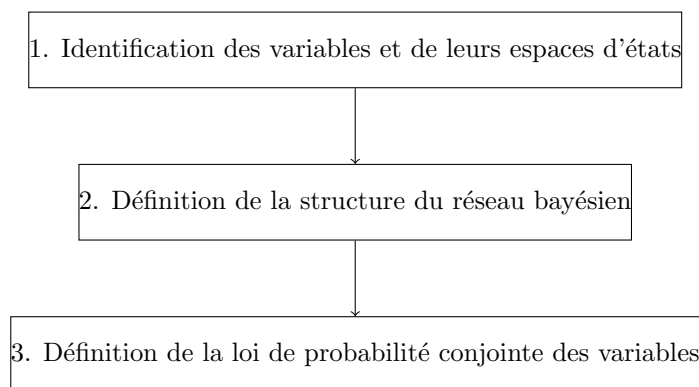
#### 4.4.4 Equivalence de markov

Soit  $G_1 = (V, E_1)$  et  $G_2 = (V, E_2)$  deux GOSCs contenant le même ensemble de nœud  $V$ ,  $G_1$  et  $G_2$  sont dites équivalents au sens de *markov*, si pour chaque trois sous-ensemble mutuellement disjoints  $A, B, C \subseteq V$ ,  $A$  et  $B$  sont D-séparés par  $C$  dans  $G_1$  *ssi*  $A$  et  $B$  sont D-séparés par  $C$  dans  $G_2$ . [3]

## 4.5 Construction d'un réseau bayésien

La construction d'un réseau bayésien s'effectue en trois étapes, Figure 1.6 :

FIGURE 1.6 – Phase de construction d'un réseau bayésien



La première étape consiste à la définition de l'ensemble des variables du système, avec précision de l'espace d'états de chaque variable. Dans le cas de grandeurs physiques, les variables seront définies sur une plage d'observations probables donnée par l'expert. Il est à noter que la plupart des logiciels de réseaux

bayésiens fonctionnent avec des variables discrètes : les variables continues devront donc être discrétisées pour un optimum entre précision des observations et saturation de l'espace mémoire machine dans le traitement du réseau.

#### 4.5.1 Discrétisation

Parmi ces méthodes on peut citer :

##### **La discrétisation par intervalles de même taille (Equal width intervals) :**

C'est une méthode très populaire à cause de sa simplicité, et généralement implémentée dans la pratique. L'algorithme de discretisation doit d'abord trier la variable en fonction de ses valeurs, puis rechercher la valeur minimale  $x_{min}$  et la valeur maximale  $x_{max}$  de cette variable. puis divise la gamme des valeurs observées pour la variable en  $k$  intervalles de même taille, où  $k$  est un paramètre fourni par l'utilisateur. La largeur de l'intervalle,  $w$ , est calculée comme suit :  $w = \frac{x_{max} - x_{min}}{k}$

##### **La discrétisation par intervalles de même fréquence (Equal frequency intervals) :**

Cette méthode est basée sur le même principe que la précédente sauf qu'au lieu de considérer des intervalles de même longueur, on considère des intervalles qui contiennent le même nombre d'individus [7].

##### **La discrétisation à une règle :**

Cette méthode est basée sur une division du domaine de chaque variable continue, à l'aide d'arbres de décision en intervalles purs (contenant les individus d'une classe particulière), chacune contenant une majorité forte d'une classe particulière sous la contrainte que chaque intervalle doit inclure au moins un certain nombre prespecifie d'individus. [8]

##### **La discrétisation basée sur le Critère d'Information d'Akaike (AIC) :**

Les auteurs de cette méthode [9] proposent une version modifiée du critère AIC afin de discrétiser l'espace des données. Cette pré-discrétisation est appliquée à la base d'exemples, elle est utilisé pour l'apprentissage d'un modèle discret.

L'AIC proposé dans [10] utilise le maximum de vraisemblance, en pénalisant les modèles comportant trop de variables, qui surapprennent les données et gé-

néralisent mal. C'est pourquoi on parle de «*vraisemblancepnalise*». Cette technique est considérée comme une approche de minimisation de la divergence (ou dissemblance). Le critère d'information AIC peut être calculé par l'équation suivante :

$$AIC = -2\log L(\hat{\theta}) + 2N \quad (1.2)$$

Où :

- $L(\hat{\theta})$  est le maximum de vraisemblance.
- $N$  est le nombre de paramètre dans le modèle.

#### 4.5.2 Apprentissage de la structure

La structure d'un réseau bayésien peut être fournie a priori par un expert, comme elle peut être extraite automatiquement à partir d'une base d'exemples. Cependant, le nombre de *GOSCs* possibles pouvant représenter  $n$  variables augmente exponentiellement avec  $n$ , comme le montre la formule récursive suivante :

$$r(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} r(n-i) = 2^{2^{O(n)}}. \quad (1.3)$$

Ce qui donne :  $r(1) = 1, r(2) = 3, r(3) = 25, r(5) = 29281, r(10) = 4,2 \times 10^{18}$ .

Cela signifie que, dans les cas où le nombre de variables dépasse *sept*, il est impossible d'effectuer une recherche exhaustive de toutes les structures dans un délai raisonnable [11]. En fait, trouver le réseau optimal à partir des données est un problème NP-difficile [12] ; Les meilleures méthodes exactes connues dans la littérature pour l'apprentissage de structure des réseaux bayésiens sont appliquées que pour les domaines où le nombre de variables est limitée (environ 30 variables). L'apprentissage de structure sera développé en détail dans le deuxième chapitre.

#### A.Fonctions d'évaluation

Généralement les méthodes d'apprentissage de structure utilisent un score permettant d'évaluer les degrés de concordance entre la structure courante et la distribution de probabilité ayant généré les données. Les méthodes basées sur les scores permettent de trouver des structures efficaces.

De nombreuses fonctions d'évaluation ont été proposées. Certaines propriétés ont été définies pour mesurer l'importance de ces métriques. Ces propriétés sont les suivantes :

**Décomposabilité :**

Les algorithmes d'apprentissage de structure utilisent différentes heuristiques dans le parcours de l'espace des structures candidates, une transition depuis une structure vers une autre se fait généralement à l'aide d'une seule opération de type ajout, soustraction ou inversion d'un arc. Cette opération permet de modifier le calcul de la probabilité jointe du domaine que sur le terme  $P(X_i | n_i)$ , où  $X_i$  est le nœud du graphe concerné par la modification de ces parents, pour éviter le recalcul du score global, le score décomposable donne la possibilité de calculer que l'impact de cette modification.

**Définition 1.7. (Score Décomposable)** Une fonction de score  $S$  est dite décomposable si, étant donné la structure  $\mathcal{G}$  d'un réseau bayésien,  $S(\mathcal{G})$  peut être exprimé sous la forme d'un produit (ou d'une somme, dans l'espace logarithmique) de scores locaux, où chaque score local dépend d'un sommet et ses parents.

$$S(\mathcal{G}) = \sum_{i=1}^n s(X_i, \pi_i) \text{ ou bien } S(\mathcal{G}) = \prod_{i=1}^n s(X_i, \pi_i)$$

où  $n$  représente le nombre de sommets du graphe  $\mathcal{G}$ .

La propriété de décomposabilité réduit considérablement la complexité du calcul du score global durant le parcours de l'espace des GOSCs.

**Équivalence :**

Deux structures sont alors dites équivalentes au sens de Markov (section 4.4.4), pour une définition complète de la notion d'équivalence) s'ils peuvent encoder une même décomposition de la loi jointe sur le domaine modélisé. Plusieurs Graphes peuvent présenter la même pertinence en terme de représentation de connaissance, avec le même ensemble d'indépendances conditionnelles.

**Définition 1.8. (Score équivalent)** Etant donné deux réseaux bayésiens  $\mathcal{G}_1$  et  $\mathcal{G}_2$  équivalents au sens de Markov, la fonction de score  $S$  est dite équivalente si, pour  $\mathcal{G}_1$  et  $\mathcal{G}_2$ ,

elle associe une même valeur.

L'utilisation d'un score équivalent permet de pouvoir distinguer des structures qui ne sont pas équivalentes (et de regrouper les structures qui le sont).

### Le score bayésien

Le score bayésien d'une structure est calculé en fonction de la vraisemblance marginale des données par rapport à cette structure.

Ce score exprime la vraisemblance d'un réseau bayésien à partir de ses paramètres.

Puisque les paramètres se calculent après la construction de la structure, on cherche la structure  $\mathcal{G}$  ayant la probabilité la plus élevée conditionnellement aux données  $D$ . c.-à-d. l'objectif principal est de maximiser la probabilité  $P(\mathcal{G}|D)$ , qui est la probabilité à posteriori.

La maximisation de cette probabilité passe au début par sa décomposition ;

Soit  $X$  l'ensemble des variables aléatoires  $\{X_1, \dots, X_n\}$ ,  $n > 1$ ,  $D$  la base de données et  $N$  le nombre d'exemples dans  $D$ , où  $\{x_{i_1}, \dots, x_{i_{r_i}}\}$ , sont les valeurs de  $X_i$ , avec  $r_i > 1, i = 1, \dots, n$ .

$\mathcal{G}$  la structure du réseau sur l'ensemble des variables  $X$ . De plus, soient  $pa_{ij}$  la  $j^{ieme}$  instantiation de  $Pa(X_i)$ , et  $N_{ijk}$  le nombre de cas dans  $D$  où  $X_i$  prend la valeur  $x_{ik}$  et que  $Pa(X_i)$  est instancié en  $pa_{ij}$ .

Le calcul de la probabilité d'une structure conditionnellement à des données se fait comme suit [27] :

$$\frac{P(\mathcal{G}_1|D)}{P(\mathcal{G}_2|D)} = \frac{\frac{P(\mathcal{G}_1, D)}{P(D)}}{\frac{P(\mathcal{G}_2, D)}{P(D)}} = \frac{P(\mathcal{G}_1, D)}{P(\mathcal{G}_2, D)} \quad (1.4)$$

Si  $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$  alors :

$$P(\mathcal{G}, D) = P(\mathcal{G})P(D|\mathcal{G}) \text{ avec } P(D|\mathcal{G}) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (1.5)$$

où  $P(\mathcal{G})$  est la probabilité a priori de la structure  $\mathcal{G}$ .

Ce résultat donne la possibilité de calculer la vraisemblance  $L(\mathcal{G}|D)$  de la structure  $\mathcal{G}$  en fonction des cas disponibles dans  $D$ .

$$L(\mathcal{G}|D) = P(D|\mathcal{G}) \quad (1.6)$$

Supposons qu'on a une distribution a priori uniforme sur l'ensemble des graphes, le rapport des deux lois jointes pour chaque structure et les données et le rapport de la distribution des données conditionnellement aux données sont les mêmes.

$$si P(\mathcal{G}_1) = P(\mathcal{G}_2) alors \frac{P(\mathcal{G}_1|D)}{P(\mathcal{G}_2|D)} = \frac{P(\mathcal{G}_1, D)}{P(\mathcal{G}_2, D)} = \frac{P(D, \mathcal{G}_1)}{P(D, \mathcal{G}_2)} \quad (1.7)$$

De cela, la variation de la vraisemblance des structures par rapport aux données est la même que la variation de la masse de la loi jointe des structures et des données, donc le score bayésien noté BD, est formulé sous la forme de la probabilité des données conditionnellement à la structure.

Le score BD est décomposable, et est défini par l'équation 1.8.

$$BD(\mathcal{G}|D) = P(D|\mathcal{G}) = \prod_{i=1}^n bd_i(X_i, Pa(X_i)|D) \quad (1.8)$$

où le score local  $bd_i$  ;

$$bd(X_i, Pa(X_i)|D) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (1.9)$$

### Le score BDe

Le score bayésien est simple à utiliser, mais son inconvénient majeur est son non équivalence (définition 1.8), Deux structures avec les mêmes indépendances conditionnelles peuvent avoir deux évaluations différentes.

Pour cela, les auteurs ont développé une nouvelle variante du score BD [28], qui se base sur l'hypothèse d'équivalence de la vraisemblance :

**Hypothèse d'équivalence de vraisemblance :** Si  $\mathcal{G}_1$  et  $\mathcal{G}_2$  sont deux structures équivalentes de probabilités a priori non négatives alors  $P(\theta|\mathcal{G}_1) = P(\theta|\mathcal{G}_2)$ .

Cette hypothèse d'équivalence de vraisemblance nécessite l'introduction de contraintes supplémentaires sur les exposants de Dirichlet  $\alpha_{ijk}$  telles que :

$$\alpha_{ijk} = N' \times P(X_i = k, Pa(X_i) = j | \mathcal{G}_c) \quad (1.10)$$

$\mathcal{G}_c$  est le graphe complètement connecté.

$N'$  est un nombre de pseudo-exemples supplémentaires défini par l'utilisateur.

Le nouveau score *BDe* est définie par l'équation 1.11.

La fonction factorielle de l'équation 1.9 est remplacé par la fonction gamma ( $\Gamma$ ) car les valeurs des  $\alpha_{ijk}$  peuvent être réels.

$$\begin{cases} P(\mathcal{G}, D) = P(\mathcal{G})P(D|\mathcal{G}) \\ avec P(D|\mathcal{G}) = P(\mathcal{G}) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})} \end{cases} \quad (1.11)$$

où  $\Gamma$  est la fonction Gamma, qui satisfait  $\Gamma(x + 1) = x\Gamma(x)$  et  $\Gamma(1) = 1$ .

Les auteurs [28] montrent que BDe est un score équivalent.

### Le score BDeu

Ce score est un cas particulier du score BDe. Ici, les distributions de probabilités a priori définies sur les paramètres du modèle évalué sont uniformes (le 'u' de BDeu signifiant uniform).

$$P(X_i = x_k, \Pi_i = \pi_{ij} | \mathcal{G}_j) = \frac{1}{r_i q_i} \quad \forall i \in 1 \dots N, k \in 1 \dots r_i, j \in 1 \dots q_i$$

$\mathcal{G}_j$  : le graphe entièrement connecté.

Les exposants de Dirichlet ont été définie dans [29, 30] comme suit :

$$\alpha_{ijk} = \frac{N'}{r_i q_i}, \quad \forall i \in 1 \dots N, k \in 1 \dots r_i, j \in 1 \dots q_i \quad (1.12)$$

où  $N'$  est un nombre aléatoire fixé.

Suivant l'équation 1.11, et pour  $N' = 1$ , le score BDeu est alors égal à :

$$S_{BDeu}(\mathcal{G}|D) = P(\mathcal{G}) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\frac{1}{q_i})}{(\Gamma(\frac{1}{q_i} + N_{ij}))} \prod_{k=1}^{r_i} \frac{(\Gamma(\frac{1}{q_i r_i} + N_{ijk}))}{\Gamma(\frac{1}{q_i r_i})} \quad (1.13)$$

La caractéristique principale du score BDeu est sa capacité de détecter les arcs correspondant à des dépendances conditionnelles de poids faibles.

Nous avons décrit quelques méthodes d'évaluations d'une structure qui se basent sur la vraisemblance du modèle évalué par rapport à la base d'apprentissage, dans ce qui suit nous allons présenter d'autres façons d'évaluations en favorisant les modèles les plus simples.

### B.Principaux scores pondérés

L'idée de base de l'utilisation d'un score pondéré est de sélectionner la meilleure structure d'un réseau bayésien qui correspond le mieux aux données, pénalisé par le nombre de paramètres nécessaires pour spécifier la distribution conjointe de probabilité.

De nombreux scores pondérés existent, tels que le Critère d'Information Bayésien (*BIC*) [31] et le critère d'information d'Akaike (*AIC*) [32]. Ces scores maximisent la probabilité des données sous le modèle avec un terme de pénalité [33] (ce terme dépend de nombre de paramètres et de la taille de l'échantillon).

Une étude faite par *Liu et al.* [34], suggère que le score *BIC* dépasse systématiquement d'autres méthodes d'évaluation telles que le critère d'information d'Akaike (*AIC*), le score d'équivalence bayésien Dirichlet (*BDeu*) et le Maximum de Vraisemblance Normalisé et factorisé (*fNML*). Il est largement utilisé dans la littérature [35,36], grâce à sa consistance et à sa simplicité.

Dans ce qui suit, nous décrivons le score pondéré utilisé dans notre travail, puis nous présentons quelques propriétés qui permettent de supprimer certains jeux de parents qui ne peuvent pas exister dans le graphique optimal.

Soit  $x_i$  un état générique de  $X_i$ , qui a un espace d'état  $\Omega_{X_i}$ . Soit  $r_i$  et  $(q_i = \prod_{X_u \in PA_i} r_u)$  les cardinalités de  $\Omega_{X_i}$  (respectivement de  $\Omega_{PA_i}$ ).

Le vecteur des paramètres est représenté par :  $\theta = \theta_{ijk}$ , tel que :

$$\theta_{ijk} = p(x_i^k | PA_i^j) \quad (1.14)$$

Avec :

- $i \in \{1, \dots, n\}$ .
- $j \in \{1, \dots, q_i\}$ .
- $k \in \{1, \dots, r_i\}$ .
- $x_i^k \in \Omega_{X_i}$ .
- $PA_i^j \in \Omega_{PA_i}$ .

Étant donné un échantillon de données  $D = \{D_1, \dots, D_N\}$  avec  $N$  instances, avec  $D_t = \{x_{v_1}^{k_1}, \dots, x_{v_n}^{k_n}\}$  représente une instance avec  $n$  valeurs, et chaque valeur  $x_{v_i}^{k_i} \in \Omega_{X_i}$ .

L'objectif principal de l'apprentissage de structure est de maximiser une fonction de score, qui peut être écrite comme :

$$\max_G S(G) = \max_G \sum_{i=1}^n S_i(PA_i) \quad (1.15)$$

Le score local  $S_i(PA_i)$  est donné comme suit :

$$S_i(PA_i) = L_{PA_i}(\theta_i) - t_i(PA_i).W \quad (1.16)$$

Où le premier terme est la vraisemblance de modèle évalué. Le deuxième est le terme de pénalité.

Avec :

- $\theta_i$  est le vecteur de paramètre associé à  $X_i$ .
- $L_{PA_i}(\theta_i) = \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} n_{ijk} \log \theta_{ijk}$ . avec  $n_{ijk}$  est le nombre d'instances dans  $D$  qui contiennent à la fois  $x_i^k$  et  $pa_i^j$ .
- $t_i(PA_i) = q_i(r_i - 1)$ .
- $W$  est un critère ( $W = \frac{\log N}{2}$  pour  $BIC$  et  $W = 1$  pour  $AIC$ ).

Les propriétés suivantes [37] nous aident à ignorer certains ensembles de parents qui ne peuvent pas exister dans le GOSC optimal.

**Lemme 1.1.** Soit  $X_i$  un nœud d'un GOSC  $G'$  pour un RB, où  $PA_i = J'$ . Supposons que  $J \subset J'$  est tel que  $S_i(J) > S_i(J')$ . Alors  $J'$  n'est pas un ensemble parent de  $X_i$  dans le GOSC optimal.

**Théorème 1.2.** En utilisant BIC ou AIC comme fonction de score et supposant que  $N \geq 4$ , soit  $G$  et  $G'$  deux GOSCs, de sorte que  $G$  est un sous-graphe de  $G'$ . Si  $G$  est tel que  $\prod_{j \in PA_i} r_j \geq N$  pour  $X_i$ , et que  $X_i$  a un sur-ensemble de parents dans  $G'$  par rapport à  $G$ , alors  $G'$  n'est pas une structure optimale.

**Corollaire 1.3.** Dans la structure optimale  $G$ , chaque nœud a au plus  $O(\log N)$  parents.

**Lemme 1.4.** Soit  $X_i$  un nœud, tel que  $J$  et  $J'$  sont deux ensembles de parents possibles, avec  $J \subset J'$  et  $t_i(J') + S_i(J) > 0$ . Alors  $J'$  et tous les sur-ensembles  $J'' \supset J'$  ne sont pas des parents optimales pour  $X_i$ .

### 4.5.3 Apprentissage des paramètres

Cette procédure permet d'estimer les distributions de probabilités conditionnelles de réseau (ou les paramètres des lois correspondantes). L'apprentissage des paramètres d'un réseau bayésien se fait à partir des données relatives au problème à modéliser. Ces données peuvent être complètes ou incomplètes. Nous nous intéressons aux problèmes avec des données complètes.

Un RB est constitué d'un graphe acyclique dirigé (GOSC)  $G = (V, E)$  (avec  $V = \{X_1, X_2, X_3, \dots, X_n\}$  correspondent à un ensemble de variables aléatoires et les arcs  $E$  représentent les dépendances directes entre ces variables), ainsi qu'un ensemble de distributions de probabilité associées à chaque variable. Pour les variables discrètes, la distribution de probabilité est décrite par un tableau de probabilité conditionnelle (TPC) qui contient la probabilité de chaque valeur de la variable étant donnée chaque instantiation de ses parents en  $G$ . Ceci est écrit sous la forme  $P(X_i|pa_i)$  où  $pa_i$  désigne l'ensemble de parents de la variable  $X_i$  dans le graphe  $G$ .

Soit  $\theta$  l'ensemble des paramètres des variables aléatoires qualitatives dans un graphe  $G$ . Soit :

- $r_i$  : la cardinalité de l'espace de  $X_i$ ,
- $|pa_i|$  : représente la cardinalité de l'espace des configurations des parents de  $X_i$ ,

- $P(X_i|pa_i = p)$  : la distribution de probabilité discrète de  $X_i$  étant donnée la configuration de  $pa_i$  état de ses parents ( $pa_i = p$ ),
- La  $k$ -ième probabilité de  $P(X_i|i = p)$  peut être représentée par  $\theta_{ipk}$ , où  $\theta_{ipk} \in \theta$ ,  $1 \leq i \leq n$ ,  $1 \leq p \leq |pa_i|$  et  $1 \leq k \leq r_i$
- $D = \{D_1, D_2, \dots, D_N\}$  est un ensemble de données entièrement observables, avec  $D_l$  est le  $l^{ième}$  exemple de  $D$ , qui est un vecteur des valeurs de toutes les variables dans  $V$

L'estimation du maximum de vraisemblance classique (MLE) [105] consiste à trouver l'ensemble des paramètres qui maximisent le logarithme de vraisemblance  $l(\theta|D) = \sum_l \log P(D_l|\theta)$ .

Soit  $N_{ipk}$  le nombre d'exemples dans  $D$  pour lesquels  $X_i$  prend sa  $k^{ième}$  valeur et ses parents  $pa_i$  prend la  $p^{ième}$  configuration. Alors  $l(\theta|D)$  peut être écrit sous la forme :

$$l(\theta|D) = \sum_{ipk} N_{ipk} \log \theta_{ipk} \quad (1.17)$$

MLE estime  $\theta$  en maximisant  $l(\theta|D)$ . En particulier, on peut obtenir l'estimation de chaque paramètre comme suit :

$$\theta_{ipk}^* = \frac{N_{ipk}}{N_{ip}} \quad (1.18)$$

Avec  $N_{ip} = \sum_{k=1}^{r_i} N_{ipk}$

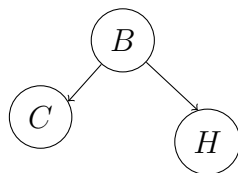


FIGURE 1.7 – Cancer métastatique

Dans cet exemple 1.7, les réseaux Bayésiens sont utilisés dans le domaine médical (Cancer métastatique), ce modèle est représenté par un GOSC, où toutes les variables(nœuds) sont Booléennes.

Nous montrons dans cet exemple comment extraire la Table de Probabilité Conditionnelle(TPC) de la variable grave mal de tête (H), qui peut être provoquée

par Les tumeurs cérébrales (B).

TABLE 1.3 – Base d'exemple du nœud **H**

B	H
Vrai	Vrai
Vrai	Faux
Faux	Faux
Vrai	Vrai
Vrai	Vrai
Faux	Faux

La Table de probabilité conditionnelle de la variable  $H$  (Tableau 1.4) est obtenu après les calculs faits en se basant sur de la base d'exemple :

- $P(H=V \mid B=V)=3/4$
- $P(H=F \mid B=V)=1/4$
- $P(H=F \mid B=F)=1$
- $P(H=V \mid B=F)=0$

L'inférence bayésienne se base sur les TPCs des différents variables pour déterminer les probabilités correspondantes aux différents événements.

TABLE 1.4 – TPC pour la variable **H**

		<b>B</b>	
		Vrai	Faux
<b>H</b>	Vrai	0.75	0
	Faux	0.25	1

## 5 La Méthode Branch & Bound ( $B\&B$ )

### 5.1 Introduction

La méthode  $B\&B$  est utilisé pour la construction d'un Réseau Bayésien. c'est une technique qui a été largement utilisée dans la résolution des problèmes d'optimisation combinatoire. Plusieurs applications pouvant être modélisées sous la forme d'un Problème d'Optimisation Combinatoire (POC) telles que le problème

du sac à dos, le problème du voyageur de commerce (PVC), le problème de la coloration de graphes, etc...

Ces problèmes contiennent des ensembles finis de solutions, où chaque solution doit être évaluée par une fonction objective, la solution doit aussi remplir certaines contraintes relatives à la nature de problème. La meilleure solution est celle dont la valeur de la fonction objective est la plus grande (resp. petite) dans le cas de maximisation (resp. minimisation) parmi l'ensemble de solutions.

Dans le cas où l'ensemble des solutions est fini, il est possible d'énumérer toutes ces solutions, et ensuite choisir la meilleure suivant la fonction objective. L'inconvénient majeur de cette approche est le nombre exponentielle des solutions possibles : il n'est guère évident de faire cette énumération.

La méthode *B&B* est une méthode d'optimisation exacte qui adopte une énumération intelligente de l'espace des solutions. Cette technique suit une stratégie permettant d'éliminer des solutions partielles qui ne mènent pas à la solution optimale. Par conséquent, cette méthode permet d'obtenir la solution recherchée en un temps raisonnable. Pour ce faire, cet algorithme utilise une fonction qui permet de mettre une borne sur certaines solutions pour soit les éliminer soit les maintenir comme des solutions potentielles. La performance d'un algorithme *B&B* dépend principalement de la pertinence de cette fonction (de sa capacité d'éliminer des solutions partielles tôt).

## 5.2 Algorithme général

L'exécution de la méthode de *B&B* peut être schématisée à travers une arborescence. Où, la racine représente l'ensemble de toutes les solutions du problème considéré.

La méthode *B&B* doit avoir :

1. une fonction de calcul d'une borne inférieure pour chaque solution partielle.
2. une stratégie de subdiviser l'espace de recherche en d'autres espaces plus

en plus petits.

3. un moyen de calcul d'une borne supérieure pour au moins une solution.

Au début, la méthode considère le problème avec son ensemble de solutions, appelé la racine.

Des bornes inférieures et supérieures sont appliquées à la racine. En cas d'égalité de ces deux bornes, la solution optimale est trouvée, et on arrête. Sinon, l'ensemble des solutions partielles sont divisées en sous-problèmes (considéré des enfants de la racine). La même procédure est appliquée à ces sous-problèmes. En cas où, une solution réalisable est obtenu, la solution optimale trouvé jusqu'à présent est modifiée si nécessaire. En cas de minimisation, si la borne inférieure d'un nœud dépasse la valeur d'une solution déjà connue (Respectivement en cas de minimisation, si la borne supérieure d'un nœud est inférieure à la valeur d'une solution déjà connue), ce nœud est éliminé, car on peut affirmer que la solution optimale globale ne peut être contenue dans le sous-ensemble de solution représenté par ce nœud.

La recherche continue jusqu'à ce que tous les nœuds sont soit explorés soit éliminés.

### 5.3 Problème de voyageur en commerce (PVC)

La manière la plus efficace de comprendre cette méthode est de l'illustrer par un exemple. Cette partie est destinée à la résolution du problème de PVC par la méthode B&B.

Soit un graphe  $G=(V,E)$ . Un cycle est dit hamiltonien *ssi* tous les sommets de  $G$  apparaissent une et une seule fois dans ce cycle.

Le problème de PVC consiste à chercher un cycle hamiltonien dont la somme des poids est minimale(Figure 1.8).

Le graphe hamiltonien peut être représenté par sa matrice d'adjacence  $D$ , (Tableau 1.5) tel que  $D[i,j]$  représente le poids de l'arc  $(i,j)$ .

FIGURE 1.8 – Exemple de PVC.

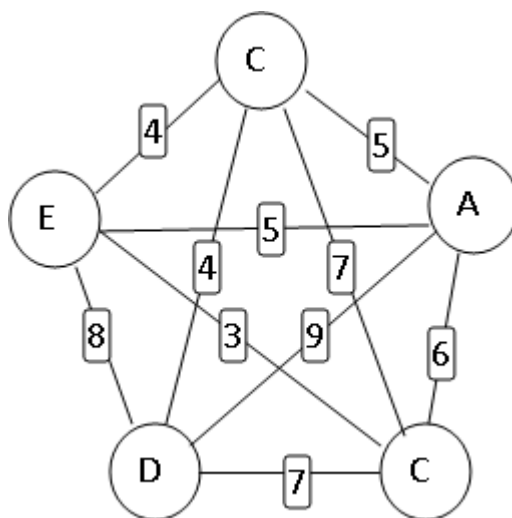


TABLE 1.5 – Matrice d'adjacence représentant les poids.

	A	B	C	D	E
A	0	6	5	9	5
B	6	0	7	7	3
C	5	7	0	4	4
D	9	7	4	0	8
E	5	3	4	8	0

Commençant à partir du sommet E, pour la recherche d'un cycle hamiltonien optimal.

Le prochain sommet du cycle est soit A, B, C ou D. Pour chacune de ces solutions partielles, nous cherchons une borne inférieure (toutes les solutions résultantes de cette solution partielle va avoir un coût  $\geq$  à cette nouvelle borne).

Initialement, un très grand nombre ou bien la meilleure solution trouvée avec une autre méthode rapide est considéré comme borne supérieure.

Le principe de la méthode est le suivant :

- Toute solution partielle dont le coût est plus grand que celui de la borne supérieure trouvée jusqu'à présent va être éliminer de la recherche.
- Si le coût de cette solution complète est inférieur à celui de la meilleure

solution que nous avons déjà, alors nous remplaçons ce coût comme étant celui de la meilleure solution courante (borne supérieure).

- Ce processus est continué jusqu'à avoir une solution optimale.

Parmi plusieurs solutions partielles, Il existe différentes manières de décider quelle solution à explorer en premier. Dans notre exemple nous choisissons toujours la solution ayant la plus petite borne inférieure.

Il faut noter que la performance d'un algorithme B&B dépend principalement de la pertinence de la fonction utilisée pour le calcul de la borne inférieure des solutions partielles.

Dans cet exemple la fonction utilisée suit le lemme suivant :

**Lemme 1.5.** *Soit le cycle hamiltonien suivant :  $v_1, v_2, \dots, v_n, v_{n+1} = v_1$ , il est facile de voir que quelque soit la solution, son coût est  $\geq \sum_{i=1}^n (\text{arc} - v_{i1} + \text{arc} - v_{i2})$  où  $\text{arc} - v_{i1}$  et  $\text{arc} - v_{i2}$  désignent deux arcs adjacents au sommet  $i$  ayant le plus petit poids.*

En commençant à partir de E, on aura comme borne inférieure :

$$\text{Borne}(E) = \frac{1}{2}\{(3 + 4) + (4 + 4) + (5 + 5) + (3 + 6) + (4 + 7)\} = 22.5$$

Les prochains sommets dans le cycle peuvent être : A, B, C ou D. Pour chacun solution partielle, une borne est calculée :

$$\text{Borne}(A) = \frac{1}{2}\{(4 + 5) + (4 + 4) + (4 + 5) + (3 + 6) + (4 + 7)\} = 23$$

$$\text{Borne}(B) = \frac{1}{2}\{(3 + 4) + (4 + 4) + (5 + 5) + (3 + 6) + (4 + 7)\} = 22.5$$

$$\text{Borne}(C) = \frac{1}{2}\{(3 + 4) + (4 + 4) + (5 + 5) + (3 + 6) + (4 + 7)\} = 22.5$$

$$\text{Borne}(D) = \frac{1}{2}\{(8 + 3) + (4 + 4) + (5 + 5) + (3 + 6) + (4 + 8)\} = 25$$

La fonction utilisé pour calculer la borne prend en compte le sommet E, c.-à-d. l'arc (E,D) doit être pris, et pour le sommet D, l'arc (D,E) doit être pris. La valeur 25 représente donc la plus petite valeur d'un cycle hamiltonien, (les arcs (E,D) et (D,E) sont inclus).



## 5.4 Parallèle Branch&Bound

Les algorithmes B&B peuvent réduire considérablement la puissance de calcul nécessaire pour explorer tout l'espace de la solution.

Toutefois, cette puissance peut être insuffisante, en particulier lors de la résolution des problèmes de grande tailles.

L'utilisation de plusieurs processeurs ou cœurs en parallèle est un moyen efficace pour réduire le temps d'exploration.

De nombreuses approches pour paralléliser des algorithmes B&B sont proposées dans la littérature. Cette section donne une brève classification de ces méthodes.

Une taxonomie des différents modèles utilisés pour paralléliser l'algorithme B&B est présentée dans [102]. Cette taxonomie est basée sur les classifications proposées dans [100] et [101]. Trois modèles peuvent être identifiés :

- Le modèle parallèle multi-paramétrique,
- Le modèle d'évaluation parallèle des bornes,
- Et le modèle d'exploration d'arbres parallèles.

Les quatre sous-sections suivantes présentent ces modèles.

### 5.4.1 Modèle parallèle multi-paramétrique

Le modèle parallèle multi-paramétrique, comme illustré dans la figure 1.10 est relativement moins étudié dans la littérature ; il est basé sur l'utilisation de plusieurs algorithmes B&B exécutés en parallèle. Plusieurs variantes de ce modèle

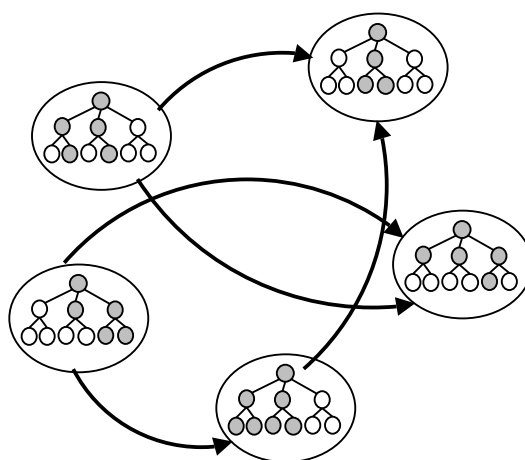


FIGURE 1.10 – Illustration du Modèle parallèle multi-paramétrique

peuvent être envisagées en fonction du choix d'un ou plusieurs paramètre(s) de l'algorithme B&B.

Les algorithmes parallèles B&B diffèrent par l'opérateur de branchement [103], comme ils peuvent avoir différents opérateurs de sélection [101] dans le cas où la stratégie d'exploration en profondeur d'abord est utilisée.

Le principal avantage du modèle parallèle multi-paramétrique est sa généralité qui permet son utilisation de manière transparente pour l'utilisateur.

Son inconvénient est la surcharge des sous-problèmes qu'il génère puisque certains sous-problèmes de l'arbre sont explorés de manière redondante. Cependant, cette surcharge supplémentaire a moins de conséquences lorsque le modèle est déployé sur un système informatique hautes performances.

#### 5.4.2 Évaluation parallèle du modèle des bornes

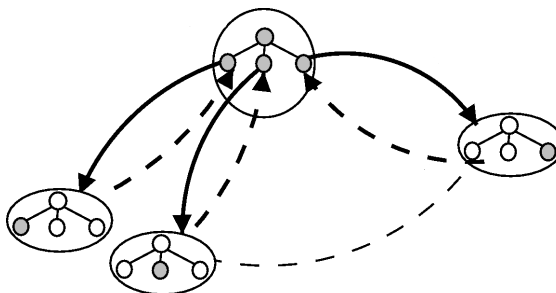


FIGURE 1.11 – Illustration de l'Évaluation parallèle du modèle des bornes

Comme le montre la figure 1.11, le modèle d'évaluation parallèle des bornes permet la parallélisation du bornage des sous-problèmes générés par l'opérateur de branchement. Ce modèle est utilisé dans le cas où l'opérateur de bornage est exécuté plusieurs fois après l'opérateur de branchement. Le modèle ne modifie pas l'ordre et le nombre de sous-problèmes explorés dans l'algorithme parallèle B&B par rapport à l'algorithme séquentiel B&B. Il faut noter que la phase de bornage est plus rapide dans l'évaluation parallèle B&B que dans le modèle séquentiel B&B, le principal avantage de ce modèle est sa généralité. Toutefois, ce modèle peut s'avérer inefficace dans certains environnements parallèles pour les raisons suivantes :

- Le modèle est synchrone et ne convient donc pas aux contextes hétérogènes et volatiles.
- Sa granularité (le coût de l'opérateur englobant) peut être fine et donc inadaptée aux systèmes de hautes performances.
- Le degré de parallélisme de ce modèle dépend du problème résolu.

La combinaison de ce modèle avec le modèle d'exploration d'arbres parallèles peut générer un degré de parallélisme plus élevé que l'utilisation de ce modèle seul.

### 5.4.3 Modèle d'exploration d'arborescence parallèle

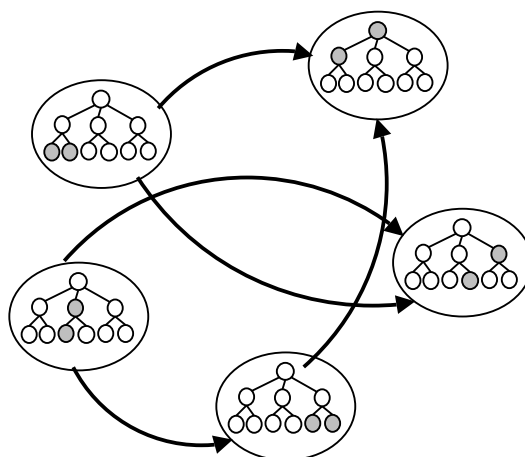


FIGURE 1.12 – Illustration du modèle d'exploration d'arborescence parallèle.

Le modèle d'exploration d'arborescence parallèle consiste à explorer simultanément plusieurs sous-problèmes des sous-arbres définissant différents sous-espaces de recherche du problème initial (Figure 1.12). Cela signifie que les opérateurs de sélection, de branchement, de bornage et d'élagage sont exécutés en parallèle de manière synchrone ou asynchrone par différents processus explorant ces sous-espaces. En mode synchrone, un algorithme B&B passe par différentes phases. Au cours de chaque phase, les processus B&B de l'algorithme effectuent leur exploration indépendamment. Entre les phases, les processus B&B sont synchronisés pour échanger des informations, telles que la meilleure solution trouvée jusqu'à présent. En mode asynchrone, les processus B&B communiquent de manière imprévisible.

Comparé à d'autres modèles, le modèle d'exploration d'arbres parallèles est

plus fréquemment utilisé et fait l'objet de nombreuses recherches pour deux raisons principales. D'une part, le degré de parallélisme de ce modèle peut être important dans les problèmes complexes avec l'utilisation d'un calcul multi-cœur ou d'un système informatique à hautes performances. D'autre part, la mise en œuvre du modèle soulève plusieurs problèmes qui constituent des défis de recherche intéressants en informatique parallèle. Parmi ces problèmes, nous pouvons inclure le placement et la gestion de la liste des sous-problèmes à résoudre, la répartition et le partage de la charge (sous-problèmes générés), la communication de la meilleure solution trouvée jusqu'à présent, la détection de la terminaison de l'algorithme, et la tolérance aux pannes.

## 6 Conclusion

Dans ce chapitre nous avons décrit l'incertitude et ses typologies, ainsi que la décision médicale qui se base sur des données prise des sources différentes, qui sont généralement incertaines. Après nous avons présenté les réseaux Bayésiens, ses propriétés et les étapes de construction de ce modèle. Finalement, la méthode *B&B* que nous avons adopté pour la construction de structure de Réseau Bayésien a été expliquée, avec une brève description des différentes stratégies permettant sa parallélisme.

Nous nous concentrons dans notre travail sur l'apprentissage de structure des réseaux bayésiens, qui représente la tâche prépondérante et la plus ardue.

Le chapitre suivant est destiné aux différentes méthodes utilisées pour la construction de ce modèle.

# Chapitre 2

## Apprentissage de Structure

### 1 Introduction

Dans la littérature, ils existent plusieurs méthodes pour l'ASRB, et peuvent être réparties en deux grandes classes :

- Les méthodes approximatives ;
- Les méthodes exactes,

Les méthodes appartenant à la première classe produisent une solution réalisable dans un temps raisonnable, et les autres cherchent la solution optimale et peuvent prendre un temps exponentiel.

### 2 Méthodes approximatives

Deux types de méthodes approximatives existent :

- Les heuristiques,
- Les métaheuristiques,

#### 2.1 Heuristiques

Dans le domaine de l'ASRB, ces méthodes peuvent être divisés en deux familles, la première est basée sur la recherche des différentes indépendances conditionnelles qui existent entre les variables du modèle, la deuxième famille tente d'associer un score à chaque réseau bayésien pour quantifier l'adéquation d'une structure au problème à résoudre, puis elle recherche le réseau bayésien qui possède le meilleur score dans l'espace des graphes acycliques dirigés.

### 2.1.1 Algorithmes d'indépendance conditionnelle

Ces algorithmes sont proposés initialement par *Sprites et al.* et *Pearl et Verma*, ils consistent à déterminer dans un premier temps un graphe non orienté en tenant compte les différentes indépendances conditionnelles qui existent entre les variables de ce graphe, puis à orienter ce graphe pour obtenir un réseau bayésien [13–17].

Il faut noter que la détermination des indépendances entre les variables devient difficile dans le cas de problèmes de grande taille.

#### 1 - Tests d'indépendance conditionnelle :

Les tests du  $\chi^2$  sont utilisés pour tester l'indépendance conditionnelle entre les variables dans un réseau bayésien. Dans ce qui suit nous allons détailler les tests d'indépendances du  $\chi^2$ , puis l'utiliser dans le cadre de l'indépendance conditionnelle.

Soit deux variables discrètes  $X_1$  et  $X_2$ , de taille respective  $r_1$  et  $r_2$ . Soit :

- $N_{12}$  le nombre d'occurrences de  $X_1 = x_1$  et  $X_2 = x_2$  dans une base d'exemples  $D$ ,
- $N_1$ . le nombre d'occurrences de  $X_1 = x_1$ ,
- $N_2$ . le nombre d'occurrences de  $X_2 = x_2$ .

Le test du  $\chi^2$  met en concurrence deux modèles :

- Le modèle observé  $p_o = P(X_1, X_2)$ , représenté par les occurrences observées  $O_{12} = N_{12}$
- Le modèle théorique  $p_t = P(X_1)P(X_2)$ , représenté par les occurrences théoriques  $T_{12} = \frac{N_1 N_2}{N}$

Soit la statistique suivante (de degré de liberté  $df = (r_1 - 1)(r_2 - 1)$ ) :

$$X^2 = \sum_{a=1}^{r_1} \sum_{b=1}^{r_2} \frac{(O_{12} - T_{12})^2}{T_{12}} = \sum_{a=1}^{r_1} \sum_{b=1}^{r_2} \frac{(N_{12} - \frac{N_1 * N_2}{N})^2}{\frac{N_1 * N_2}{N}}$$

$X_1$  et  $X_2$  sont indépendant si et seulement si  $\chi^2 < \chi^2$  théorique(df, 1-  $\alpha$ )

avec  $\alpha$  un seuil de confiance qui varie entre 0 et 1.

*Algorithme PC* Plusieurs algorithmes génèrent la structure de réseau bayésien en se basant sur la détermination des indépendances conditionnelles à partir de données.

Dans ce qui suit, nous décrivons en détail l'algorithme PC (d'après ses auteurs, Peter et Clark) [17], qui est utilisé pour l'apprentissage de la structure d'un réseau bayésien en se basant sur le test de  $\chi^2$  (voir l'algorithme **PC**).

L'algorithme suit ces étapes :

- la construction d'un graphe non orienté contenant les relations entre les variables, en utilisant les tests d'indépendance conditionnelle,
- la détection des V-structures (les tests d'indépendance conditionnelle sont aussi utilisés dans cette étape),
- la propagation des orientations de certains arcs.

L'algorithme **PC** fonctionne comme suit :

Dans un graphe  $G(V, E, \theta)$ , deux sommets  $X_i$  et  $X_j$  de  $V$  et un sous ensemble de sommets  $S_{X_i, X_j} \in V \setminus \{X_i, X_j\}$ ,  $X_i$  et  $X_j$  sont reliés par un arc dans  $G$  s'il n'existe pas  $S_{X_i, X_j}$  tel que  $(X_i \perp\!\!\!\perp X_j | S_{X_i, X_j})$

Le test de  $\chi^2$  est utilisé pour la vérification de l'existence d'une telle indépendance  $(X_i \perp\!\!\!\perp X_j | S_{X_i, X_j})$  par ordre croissant, à partir d'un graphe non orienté entièrement connecté.

La détection d'indépendances permet alors de supprimer les arêtes correspondantes, cette opération se répète jusqu'à l'obtention du squelette du GOSC recherché.

Cet algorithme suit les deux phases suivantes :

- Détecter et établir les V-structures du graphe ;
- Orienter les arêtes restantes .

Les règles d'orientation des arêtes restantes du graphe obtenu à l'issue de la phase de détection, peuvent être remplacées par une heuristique.

La deuxième phase consiste à orienter le graphe de manière à :

- Ne pas créer de circuit ;
- De ne pas créer de V-structure ;

En parallèle à l'algorithme PC, il existe un autre algorithme, nommé IC (Inductive Causation), présenté par *Pearl* [16]. Cet algorithme est similaire à l'algorithme PC, mais il commence par une structure vide, et relie les couples de variables lorsqu'une dépendance conditionnelle est détectée.

Il existe une autre variante de l'algorithme IC, nommé IC\*, cet algorithme prend en compte les variables latentes [18].

### 2.1.2 Algorithmes basés sur les scores

Cette famille d'algorithmes consiste à associer un score à chaque graphe candidat, ensuite choisir le graphe ayant le score le plus élevé.

Ces méthodes sont compliquées, à cause de la taille super exponentielle de l'espace de recherche en fonction du nombre de variables.

Dans ce qui suit nous allons décrire quelques méthodes appartenant à cette famille :

#### *1-Arbre de Recouvrement de Poids Minimal (ARPM) :*

cette méthode est utilisée pour la recherche de l'arbre de recouvrement de poids minimal. Elle s'applique aussi à l'ASRB. Au début, un poids est fixé à chaque arête potentielle de l'arbre, ce poids peut être par exemple l'information mutuelle entre les variables. Après la définition de la matrice de poids, l'ARPM utilise un algorithme pour la résolution du problème de l'arbre de poids minimum comme l'algorithme de *Kruskal* ou celui de *Prim*.

Le résultat obtenu par cet algorithme est un arbre non dirigé (voir l'algorithme **ARPM**), qui doit être orienté en choisissant une racine puis en parcourant l'arbre par une recherche en profondeur [19].

La racine du départ peut être choisie aléatoirement ou à l'aide de connaissance a priori, ou encore avec la variable représentant la classe dans le cas des problèmes de classification.

L'efficacité de ces méthodes dépend principalement de la pertinence des heuristiques utilisées pour l'exploration de l'espace des graphes.

### 2- Recherche Gloutonne (RG)

L'algorithme classique de RG est utilisé par *Chickering et al.* [20] pour l'ASRB, où trois opérateurs sont utilisés pour définir le voisinage d'un graphe (voir algorithme **RG**) : *ajout*, *suppression* ou *inversion d'arc*, les scores locaux sont utilisés pour simplifier le calcul du score global.

Pour cet exemple de petite taille (figure 2.1), le voisinage du graphe courant contient neuf GOSCs, donc le graphe ayant la meilleure qualité (score) est choisi pour la prochaine itération.

Le calcul de score devient important pour des structures plus complexes, donc l'utilisation des scores locaux pour limiter les calculs devient une nécessité, ainsi qu'un système de cache est considéré pour éviter le recalcul du même score local plusieurs fois.

### 3- Algorithme K2

L'évaluation des différents GOSC de l'algorithme K2 [21] est réalisée à l'aide d'un score. En raison de la taille exponentielle de l'espace GOSC par rapport au nombre de variables, cet algorithme réduit l'espace de recherche avec deux méthodes (voir l'algorithme **K2**) :

- Il prend en entrée un ordre d'énumération sur les variables du domaine,
- Il limite le nombre des parents des variables.

L'ajout des parents se fait selon un ordre fixé au début, c.-à-d. le premier nœud ne peut pas posséder de parents. Il choisit toujours l'ensemble des parents qui obtient le meilleur score bayésien.

## 2.2 Métaheuristiques

Dans ce qui suit, nous allons citer quelques métaheuristiques utilisées pour l'ASRB :

Dans l'article [69], les auteurs ont proposé une métaheuristique appelée I-

ACO-B ; qui améliore la méthode de la colonie de fourmis pour l'ASRB [94] (ACO-B). Elle est basée sur le test d'indépendance conditionnelle. Premièrement, l'ACO-B utilise des tests d'indépendance d'ordre 0 pour restreindre efficacement l'espace des solutions candidates, de manière à éviter de nombreuses recherches inutiles des fourmis. Et dans le but de combiner l'augmentation du score global d'une solution et l'information mutuelle locale entre les nœuds, une nouvelle fonction heuristique dotée d'une meilleure capacité est utilisée pour initier le processus de recherches stochastiques. Les résultats expérimentaux montrent que le nouvel algorithme est efficace dans les bases de données volumineux, en plus, il améliore considérablement la vitesse de convergence par rapport à l'algorithme d'origine.

la méthode proposée dans ce travail est basée sur les algorithmes chaotiques de particules d'essaim d'optimisation (CPSO) et l'algorithme K2 [70]. Comme l'algorithme PSO piège complètement le minimum local dans les évolutions ultérieures, ils ont combiné l'algorithme PSO avec la théorie du chaos, qui possède les propriétés d'ergodicité, de caractère aléatoire et de régularité.

La méthode proposée utilise le CPSO pour identifier l'ordre des variables pour l'algorithme K2, ils ont aussi utilisé la carte chaotique pour mettre à jour les particules.

Les résultats expérimentaux montrent que la méthode proposée peut améliorer le taux de convergence des particules et identifier les structures des réseaux bayésiens de manière plus efficace et plus précise.

Un autre travail expose une approche nommé K2GA, pour l'ASRB à partir de données. K2GA utilise un algorithme génétique pour effectuer une recherche stochastique, tout en utilisant une version modifiée de l'heuristique K2 pour évaluer les réseaux proposés et améliorer les générations futures. Ce travail décrit chaque composant de K2GA et la combinaison de ces composants pour former l'algorithme de base pour une meilleurs précision [71].

L'ordre des variables (VO) ont été utilisées comme une restriction dans le processus d'apprentissage. Les informations VO peuvent réduire considérablement l'espace de recherche et permettre à certains algorithmes d'atteindre de bons ré-

sultats [72]. Les travaux antérieurs rapportés dans la littérature suggèrent que la combinaison des algorithmes d'évolutions (EA) et des VO est utile pour l'ASRB à partir de données. Cependant, la plupart des travaux sur ce domaine n'explorent pas les caractéristiques spécifiques du domaine. Ils appliquent donc simplement des opérateurs d'évolution classiques. De plus, la plupart des travaux ne rapportaient pas de bons résultats s'ils s'appliquaient à de grands RB. Cet article propose un nouvel opérateur de mutation, appelé opérateur de mutation basé sur la distance (DMO), à utiliser avec l'algorithme d'évolution à classement des variables (VOEA). Les résultats expérimentaux obtenus par VOEA sont comparés à ceux obtenus par VOGA (algorithme génétique à ordre variable) et indiquent une amélioration de la qualité de la structure des réseaux bayésiens.

Cet article propose un nouvel algorithme appelé ACO-E, destiné à l'ASRB. Pour ce faire, il effectue une recherche dans l'espace des classes d'équivalence des réseaux bayésiens à l'aide de l'optimisation de la colonie des fourmis (ACO). À cette fin, deux nouvelles extensions des techniques ACO traditionnelles sont proposées et mises en œuvre [73]. Premièrement, plusieurs types de mouvements sont autorisés. Deuxièmement, les déplacements peuvent être donnés en termes d'indices qui ne sont pas basés sur des nœuds de graphe de construction. Les résultats des tests montrent que ACO-E est plus performant qu'une recherche gloutonne et d'autres métaheuristiques lorsqu'il effectue une recherche dans l'espace des classes d'équivalence.

Ce travail présente une nouvelle approche pour l'ASRB [74] : les auteurs abordent le problème de la recherche de la meilleure structure de réseau bayésienne, à partir d'une base de données de cas, en utilisant la philosophie de l'algorithme génétique ; ils supposent un ordre entre les nœuds des structures de réseau. Cette hypothèse est nécessaire pour garantir que les réseaux créés par les algorithmes génétiques sont des structures de réseaux légales bayésiennes. Ensuite, ils utilisent un «*rparateur*» qui convertit les structures illégales en structures légales. Ils ont utilisé des simulations sur deux bases de données : ASIA et ALARM, ils ont aussi effectué une analyse de performance des paramètres de contrôle des algorithmes génétiques (taille de la population, optimiseur local, mécanisme de réduction, probabilité de croisement, taux de mutation). Les résultats obtenus

indiquent qu'en utilisant des algorithmes génétiques dans l'ASRB, il est recommandé d'appliquer un algorithme hybride qui utilise la réduction élitiste en combinaison avec une population pas trop petite et un taux de mutation relativement faible. Ceci est encore plus vrai si aucune restriction d'ordre n'est supposée entre les nœuds.

Les algorithmes génétiques avec une fonction de score robuste sont considérés pour l'ASBR [75]. Pour assurer la proximité des opérateurs génétiques, l'ordre des variables (nœuds) doit être déterminé. Dans cet article, les auteurs ont considéré l'analyse en composantes principales (ACP) pour déterminer l'ordre des nœuds. À cette fin, l'algorithme cherche d'abord la corrélation appropriée entre les variables, puis utilise la valeur absolue des coefficients de la variable dans la première composante. Cela signifie qu'un nœud  $X_i$  ne peut avoir le nœud  $X_j$  comme parent que si la valeur absolue du coefficient  $X_j$  dans la première composante est supérieur à celle de  $X_i$ . L'algorithme génétique est ensuite utilisé avec un ordre des nœuds pour construire le réseau bayésien. Le score  $BIC$  est utilisé pour mesurer la qualité des structures.

Les résultats expérimentaux obtenus sur des bases de données bien connus tels que *Asia*, *Alarm* et *Hailfinder* montrent que cette nouvelle technique offre des meilleurs résultats.

De plus, cette technique a été appliquée à un ensemble de données réelles concernant le débiteur de la Banque, qui doit contenir plus de 500 millions de *rials* à la *Maskan Bank* en *Iran*, les résultats montrent également que la technique proposée a un pouvoir de modélisation plus important que d'autres techniques.

### 3 Méthodes exactes

Dans le domaine de l'ASRB, Plusieurs méthodes de résolution exactes sont utilisées pour la résolution des problèmes de taille limités, (jusqu'à 30 variables) [37]. Ces méthodes sont très intéressantes, elles permettent de garantir l'optimalité des solutions trouvées, malgré que le temps d'exécution peut être exponentiel. Dans ce qui suit nous allons décrire les méthodes exactes les plus populaires pour l'ASBR :

### 3.1 Programmation dynamique

Soit un ensemble de données  $D$  complètement observé sur  $n$  variables discrètes [22], et soit un réseau bayésien  $G = (V, E)$ , où chaque nœud correspond à une variable et chaque arc représente une relation de dépendance entre deux variables. La fonction  $DagScore(D)$  représente un score décomposable :

$$DagScore(G) = \sum_{X \in V} NouedScore(X|parents(X)) \quad (2.1)$$

où  $NouedScore(X|parents(X))$  sont les paramètres de la probabilité conditionnelle de la variable  $X$  sachant ses parents.

Cette méthode utilise la notion d'ordre des variables. Un ordre est une permutation des variables, où si  $p_i = X$  alors  $parents(X) \subseteq \{p_1 \dots p_{i-1}\}$ . Tout graphe respectant cette contrainte doit être acyclique. Un ordre code un ensemble de  $O(2^n)$  possible structures, et une structure peut être compatible avec de nombreux ordres.

Cet algorithme commence avec une question très simple pour la découverte de structure complète. Chaque GOSC a au moins une feuille. Peut-on identifier une feuille du meilleur structure,  $X_f = Feuille(V)$  ?

Puisque  $X_f$  ne peut être le parent d'aucune autre variable, le score du meilleur réseau sur  $V$  est composé de deux parties : (i) le score du meilleur réseau sur  $V - \{X_f\}$  et (ii) le score de  $Feuille(V)$  étant donné le meilleur ensemble de parents à partir de  $V - \{X_f\}$ , c'est l'ensemble des parents qui maximise le score du nœud  $X_f$ . Cet argument nécessite un score décomposable.

Dans chaque appel récursif, l'algorithme demande quel est le nœud qui est une feuille dans la meilleure structure sur  $V - \{X_f\}$ . Ce processus induit un ordre (inverse) sur les variables.

Formellement cette méthode peut être décrite comme suit :

$$Score(V) = \max_{X \subseteq V} Score(V - \{X\}) + MeilScore(V, X) \quad (2.2)$$

$$Feuille(V) = \arg \max_{X \subseteq V} Score(V - \{X\}) + MeilScore(V, X) \quad (2.3)$$

$$MeilScore(V, X) = \max_{PS \subseteq V - \{X\}} NouedScore(X|PS)$$

Où  $score(\emptyset) = 0$ , et  $Feuille(\emptyset) = nil$  et  $PS$  un ensemble des parents.

Pour visualiser l'algorithme, considérons le réseau de sous-ensembles de la figure 2.2. Il y'a un état pour chaque sous-ensemble de  $V$ . Les enfants (successeurs) de l'état  $S$  dans le réseau correspondent à des choix pour la feuille  $S$ .

Une première recherche naïve en profondeur dans le graphe correspond à l'évaluation directe des récurrences 2.2 et 2.3.

L'algorithme commence au bas du graphe de la figure 2.2, et recherche tous les chemins vers le haut. Chaque transition de l'état  $S$  à un successeur correspond au choix d'une feuille dans un réseau sur les nœuds de  $S$ . Ainsi, un seul chemin de  $\{1, 2, \dots, n\}$  à  $\{\}$  correspond à une séquence de suppression de feuilles.

Une simple recherche en profondeur nécessite  $O(n!)$  en Temps et  $O(n)$  en espace. L'alternative est plus complexe en mémoire. Une fois que la valeur de  $Score(S)$  est calculée sur un chemin de recherche, elle peut être mémorisée pour être réutilisée. Si  $Score(S)$  est mis en cache, les chemins de recherche futurs qui arrivent à  $S$  peuvent être élagués. C'est une programmation dynamique sur le réseau de sous-ensembles (Voir l'algorithme **PrgDyn**).

Lorsque l'algorithme **PrgDyn** se termine, la valeur de  $Feuille(S)$  est connue pour chaque  $S \subseteq V$ . Le dernier élément de la commande est  $Feuille(V)$ . Le second dernier élément de l'ordre est  $Feuille(V - Feuille(V))$ , etc.

En résumé, l'algorithme de programmation dynamique fonctionne comme suit. Il cherche d'abord des structures optimales pour des variables uniques, ce qui est simple. En partant de ces structures de base, l'algorithme construit des sous-réseaux optimaux pour des ensembles de variables de plus en plus grands jusqu'à trouver un réseau optimal pour  $V$ .

L'inconvénient majeur de cet approche est son besoin de trouver des sous-réseaux optimaux pour tous les sous-ensembles de variables, ce qui nécessite le calcul de tous les scores des parents pour chaque variable; pour  $n$  variables, il

y a  $2^n$  nœuds dans le graphe d'ordre, et  $n$  graphes parents avec  $2^{n-1}$  nœuds parents chacun et, le nombre des scores des parents qui doit être calculés est  $n2^{n-1}$ . Par conséquent, le calcul et le stockage de l'ordre et des graphes des parents deviennent rapidement irréalisables, avec un nombre important de variables. D'après les expérimentations faites, l'application de cette méthode est limitée à 27 variables.

### 3.2 Algorithme de recherche $A^*$

L'idée de base de cet algorithme est de formuler l'apprentissage optimal d'un réseau bayésien comme un problème de recherche du plus court chemin [23]. Le graphe d'ordre de la figure 2.3 est utilisé comme graphique de recherche. Le nœud le plus haut qui ne contient pas de variables est considéré comme étant l'état de départ et le nœud le plus bas avec toutes les variables comme l'état de but. Pour deux nœuds voisins  $S1$  et  $S2$  avec  $S1$  étant parent de  $S2$ , le coût de l'arc  $c(S1, S2)$  est défini à être  $MeilMDL(X, S1)$ , où  $X$  est la seule variable dans  $S2 \setminus S1$ . L'objectif de cet algorithme est alors de trouver le chemin le plus court entre l'état de départ et l'état de but ayant le coût le plus bas. Par définition, le chemin le plus court correspond à un réseau bayésien optimal.

L' algorithme de recherche  $A^*$  utilise une file d'attente prioritaire, appelée OVERT list, pour organiser la recherche, qui est initialisé avec l'état de départ. À chaque étape de recherche, le nœud de la liste OVERT ayant le coût le plus faible est sélectionné pour le prochain développement.

pour chaque nœud résultat, le coût total estimé ( $fcout$ ) est calculé comme la somme des coûts exact jusqu'à présent ( $gcout$ ) et le coût estimé de l'état objectif ( $hcout$ ).

Une fois un nœud est développé, il est placé dans une liste  $FERME$ . La détection des doublons est effectuée pour chaque nœud nouvellement généré sur les listes  $OVERT$  et  $FERME$ . Si un doublon est détecté dans la liste  $FERME$ , le nouveau nœud est rejeté en se basant sur une heuristique. Si un doublon est détecté dans la liste  $OVERT$  et que le nouveau nœud a un coût  $gcout$  inférieur,

le nœud existant est mis à jour avec le nouveau coût  $gcost$  et ses parents. Une fois l'état objectif est atteint, la structure optimale du réseau est extraite facilement à partir du chemin complet de l'état de départ à l'état de but.

Le coût  $gcost$  d'un nœud est calculé comme la somme des coûts des arcs sur le meilleur chemin entre l'état de départ et le nœud actuel. Chaque coût de l'arc est calculé lorsqu'un successeur est généré pendant la recherche, en récupérant des informations d'un nœud correspondant dans le graphe parent. Puisque la recherche  $A^*$  explore juste une partie du graphe d'ordre, la méthode a besoin de calculer qu'une partie des coûts des arcs. Cet élagage est inhérent à l'algorithme de recherche et ne dépend d'aucune propriété de la fonction score.

Le coût  $hcost$  est calculé à l'aide d'une fonction heuristique. Si cette fonction est cohérente, l'algorithme de recherche  $A^*$  garantit de trouver le plus court chemin vers un nœud une fois ce nœud est sélectionné pour le développement. Par conséquent, le plus court chemin qui correspond à un réseau bayésien optimal est trouvé une fois l'état de but est sélectionné pour le développement.

Soit  $U$  un nœud dans le graphe d'ordre, la fonction heuristique  $h$  est défini comme suit :

$$h(U) = \sum_{X \in V \setminus U} BestMDL(X, V \setminus \{X\}). \quad (2.4)$$

La fonction heuristique  $h$  est clairement admissible, car elle permet à chaque variable restante de choisir les parents parmi toutes les autres variables de  $V$ , ce qui détend efficacement la contrainte acyclique et se traduit par un coût inférieur.

### 3.3 Programmation linéaire

Le problème d'ASRB est transformé en un programme linéaire sur un polytope  $\mathcal{P}$  de structures acycliques valides [24]. Le polytope est défini comme une coque convexe de sommets, où chaque sommet correspond à une structure acyclique valide. Les sommets sont représentés comme des vecteurs binaires où chaque coordonnée correspond à un choix possible de parents pour une variable.

Plus précisément, un graphe acyclique est représenté avec un vecteur binaire  $\eta = [\eta_1; \dots; \eta_n]$  où chaque  $\eta_i$  est un vecteur indicateur (de dimension  $|P_a(i)|$ ) spécifiant l'ensemble parent choisi pour le nœud correspondant. En d'autres termes, si le nœud  $i$  sélectionne les parents  $s_i$ , alors  $\eta_i(s_i) = 1$  et toutes les coordonnées restantes de  $\eta_i$  sont nulles. Notez aussi que  $\eta$  est un vecteur clairsemé avec exactement  $n$  coordonnées égales à un.  $\eta(s)$  est utilisé pour dénoter le vecteur binaire correspondant au graphe  $G(s)$  obtenu en sélectionnant les ensembles parents  $s = [s_1; \dots; s_n]$ .

Le polytope  $\mathcal{P}$  est maintenant la coque convexe de tout  $n(s)$  où  $s_i$  appartient à  $Pa(i)$  et  $s = [s_1; \dots; s_n]$  correspond à un GOSC. La propriété clé de ce polytope  $\mathcal{P}$  est que pour tout graphe  $G(s)$  avec cycles est garanti d'être situé à l'extérieur de  $\mathcal{P}$ .

Des moyennes pondérées de vecteurs binaires  $\eta$  ( $\eta \in \mathcal{P}$ ) ont été utilisées pour des graphes acycliques.

La formulation du problème d'ASRB peut être définie comme suit :

$$\begin{aligned} \max \quad & \eta \cdot W = \sum_{i=1}^n \sum_{s_i \in Pa(i)} \eta_i(s_i) W_i(s_i) \\ \text{s.t.} \quad & \eta \in \mathcal{P} \end{aligned} \tag{2.5}$$

avec  $W_i(s_i)$  est le score local et  $n$  est le nombre de variables.

Le réseau bayésien ayant obtenu le score le plus élevé représente la solution optimale de ce problème d'apprentissage. La complexité de cet algorithme est exponentielle.

Il faut noter que  $\mathcal{P}$  défini ci-dessus est différent du sous-graphe acyclique polytope  $\mathcal{P}_{dag}$  étudié en détail dans la combinatoire polyédrique.  $\mathcal{P}_{dag}$  est défini par *Grotschel et al.* [25] comme étant la coque convexe des vecteurs indicateurs d'arc pour chaque ensemble d'arcs spécifiant un graphe acyclique, cette contrainte n'est pas défini dans  $\mathcal{P}$  (voir l'équation 2.5).

Cependant, le score associé à un réseau bayésien dépend de l'ensemble des arcs entrants (ensembles parents). Ce score n'est pas une fonction linéaire des

arcs individuels, pour cette raison  $\mathcal{P}_{dag}$  ne serait pas capable de convertir le problème d'ASRB en un programme linéaire.

Dans [26], les auteurs ont créé une variable de famille  $X_{i \leftarrow J}$  pour chaque famille potentielle, pour formuler le problème d'ASRB. Une variable de famille est une variable d'indicateur binaire :  $X_{i \leftarrow J} = 1$  si  $J$  est l'ensemble parent pour  $i$  et  $X_{i \leftarrow J} = 0$  sinon. Il n'est pas difficile de voir qu'un digraphe (acyclique ou autre) avec  $|V|$  nœuds qui peuvent être codés par un vecteur *zero-one*, dont les composants sont des variables de famille, ou exactement  $|V|$  variables de famille sont définies sur 1.

La figure 2.5 et le tableau 2.1 montrent un exemple de graphique et du codage de variable de famille, respectivement.

Les contraintes les plus élémentaires sont illustrées par la disposition du vecteur d'exemple dans le tableau 2.1 en trois lignes, une pour chaque nœud. Il est clair qu'une seule variable de famille pour chaque nœud enfant doit être égale à un. Donc  $|V|$  représente les contraintes de convexité.

$$\sum_{J \in \mathcal{P}} X_{i \leftarrow J} = 1 \forall i \in V \quad (2.6)$$

Chaque vecteur peut encoder un nombre exponentiel de graphes.

Parmi les contraintes d'acyclicités des graphes existantes dans la littérature, les auteurs dans [24] ont proposé une méthode, où un cluster a été défini par un sous-ensemble de nœuds avec au moins 2 éléments.

Pour chaque cluster  $C \subseteq V (|C| > 1)$  l'équation de l'inégalité de cluster est formulé sous cette forme :

$$\sum_{i \in C} \sum_{J \in \mathcal{P}(i): J \cap C = \emptyset} X_{i \leftarrow J} \geq 1 \quad (2.7)$$

Pour tout cluster  $C$  et graphe  $G$ , le côté gauche de (2.7) est un décompte du nombre de sommets dans  $C$  qui est dans  $G$  et qui n'a pas de parents dans  $C$ . Supposons maintenant que les nœuds d'un cluster  $C$  forment un cycle ; il est clair que dans ce cas, le côté gauche de (l'équation 2.7) serait 0, d'où, il y'a violation de la contrainte de cluster.

La formulation du problème d'ASRB peut être définie comme suit :

$$MAX \sum_{i \in V, J \in \mathcal{P}(i)} c_{i \leftarrow J} X_{i \leftarrow J} \quad (2.8)$$

$$\sum_{J \in \mathcal{P}} X_{i \leftarrow j} = 1 \quad \forall i \in V \quad (2.9)$$

$$\sum_{i \in C} \sum_{J \in \mathcal{P}(i): J \cap C = \emptyset} X_{i \leftarrow j} \geq 1 \quad \forall C \subseteq V, |C| > 1 \quad (2.10)$$

$$X_{i \leftarrow J} \in \{0, 1\} \quad \forall i \in V, J \in \mathcal{P}(i) \quad (2.11)$$

## 4 Conclusion :

Ce chapitre décrit les différentes méthodes d'ASRB, premièrement nous avons décrit les méthodes approchés qui produisent des solutions réalisables qui ne sont pas toujours optimales, et puis les autres méthodes exactes les plus populaires dans ce domaine tel que la programmation dynamique est l'algorithme A\*. Concernant la méthode *Branch&Bound* appliqué à l'ASRB, que nous avons amélioré, elle sera décrite en détail dans le quatrième chapitre. Le chapitre suivant sera consacré aux différentes applications des réseaux bayésiens dans le domaine médical.

**Algorithme PC**

- Construction d'un graphe non orienté

Soit  $G$  le graphe reliant complètement tous les nœuds  $X$

$i \leftarrow 0$

Répéter

Recherche des indépendances cond. d'ordre  $i$

$\forall \{X_A, X_B\} \in X^2$  tels que  $X_A - X_B$  et  $\text{Card}(\text{Adj}(G, X_A, X_B)) \geq i$

$\forall S \subset \text{Adj}(G, X_A, X_B)$  tel que  $\text{Card}(S) = i$

si  $X_A \perp X_B | S$  alors :

◇ suppression de l'arête  $X_A - X_B$  dans  $G$

◇  $\text{SepSet}(X_A, X_B) \leftarrow \text{SepSet}(X_A, X_B) \cup S$

◇  $\text{SepSet}(X_B, X_A) \leftarrow \text{SepSet}(X_B, X_A) \cup S$

$i \leftarrow i + 1$

Jusqu'à  $\text{Card}(\text{Adj}(G, X_A, X_B)) < i, \forall X_A, X_B \in X^2$

- Recherche des V-structures

$\forall \{X_A, X_B, X_C\} \in X^3$  tels que  $\overline{X_A X_B}$  et  $X_A - X_C - X_B$ ,

si  $X_C \notin \text{SepSet}(X_A, X_B)$  alors on crée une V-structure :

$X_A \rightarrow X_C \leftarrow X_B$

- Ajout récursif de  $\rightarrow$

Répéter

$\forall \{X_A, X_B\} \in X^2$ ,

si  $X_A - X_B$  et  $X_A \rightsquigarrow X_B$ , alors ajout d'une flèche à  $X_B$  :

$X_A \rightarrow X_B$

si  $\overline{X_A X_B}$ ,  $\forall X_C$  tel que  $X_A \rightarrow X_C$  et  $X_C - X_B$  alors  $X_C \rightarrow X_B$

Tant qu'il est possible d'orienter des arêtes

**Notations :**

- $X$  : ensemble de tous les nœuds,
- $\text{Adj}(G, X_A)$  : ensemble des nœuds adjacents à  $X_A$  dans  $G$ ,
- $\text{Adj}(G, X_A, X_B) : \text{Adj}(G, X_A) \setminus \{X_B\}$ ,
- $X_A - X_B$  : il existe une arête entre  $X_A$  et  $X_B$ ,
- $X_A \rightarrow X_B$  : il existe un arc de  $X_A$  vers  $X_B$ ,
- $\overline{X_A X_B}$  :  $X_A$  et  $X_B$  adjacents  $X_A - X_B$ ,  $X_A \rightarrow X_B$  ou  $X_B \rightarrow X_A$ ,
- $X_A \rightsquigarrow X_B$  : il existe un chemin dirigé reliant  $X_A$  et  $X_B$ ,

---

**Algorithme ARPM « dirigé »**

- Construction de l'arbre optimal (Kruskal)

$$\forall X_i, T(X_i) = \{X_i\}$$

$$B \leftarrow \emptyset$$

$$\forall (X_i, X_j) \in A$$

si  $T(X_i) \neq T(X_j)$  alors

- $B \leftarrow B \cup (X_i, X_j)$
- $T' \leftarrow T(X_i) \cup T(X_j)$
- $T(X_i) \leftarrow T'$
- $T(X_j) \leftarrow T'$

- Orientation des arêtes

$$B \leftarrow \emptyset$$

$$\{pa_i\} \leftarrow \text{ParcoursProfondeur}(B, X_r)$$

$$\forall X_i,$$

si  $pa_i \neq \emptyset$  alors ajout de  $pa_i \rightarrow X_i$  dans  $B_f$

---

Notations :

**A** : liste des arêtes  $(X_i, X_j)$  dans l'ordre décroissant des poids,

**T** $(X_i)$  : arbre passant par le nœud  $X_i$ ,

$X_r$  : racine choisie pour orienter l'arbre,

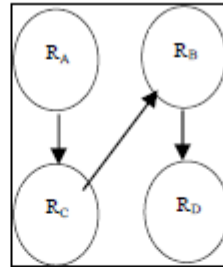
$pa_i$  : parent du nœud  $X_i$ ,

**B** : arbre optimal non orienté,

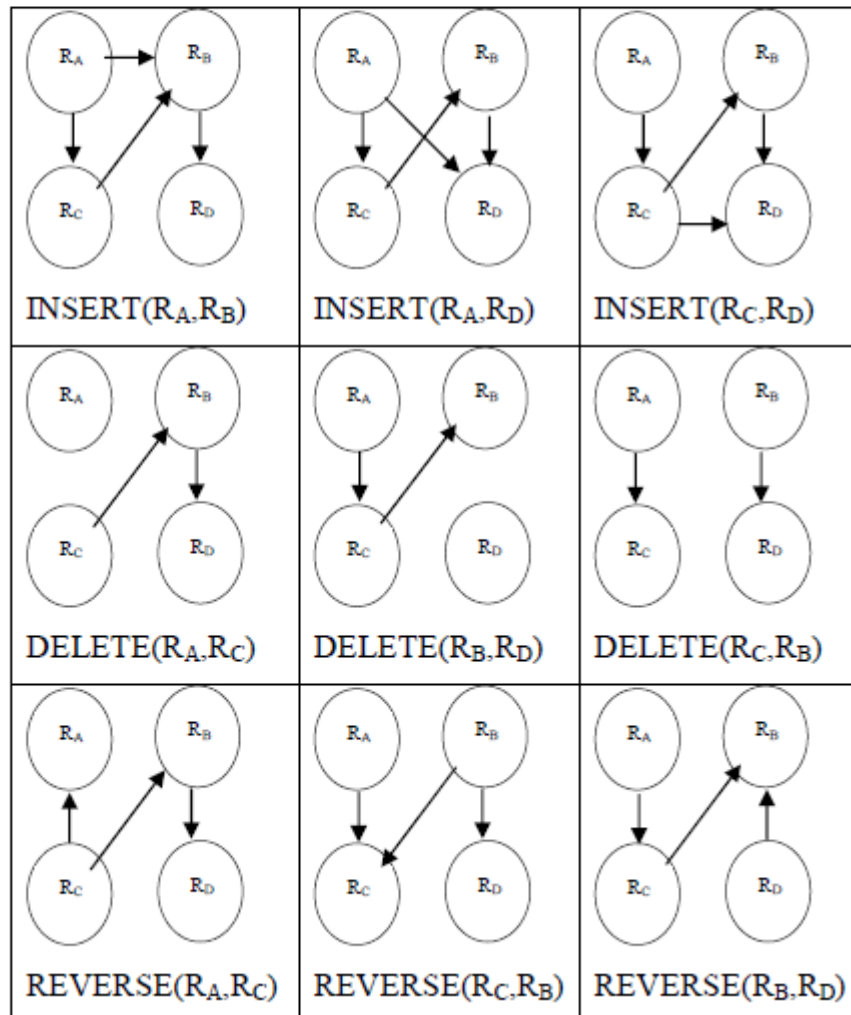
$B_f$  : structure finale obtenue par l'algorithme.

---

FIGURE 2.1 – Exemple de recherche gloutonne.



Génération du voisinage



**Algorithme RG**

- Initialisation du graphe G  
(Graphe vide, aléatoire, donné par un expert ou arbre obtenu par ARPM)
  - Continuer  $\leftarrow$  Vrai
  - $Score_{max} \leftarrow score(G)$
  - Répéter
    - Génération de  $V_G$ , voisinage de G, à l'aide d'opérateurs :
      - Ajout d'arc, suppression d'arc, inversion d'arc  
(les graphes ainsi obtenus doivent être acycliques)
    - Calcul du score pour chaque graphe de  $V_G$
    - $G_{new}, argmax_{G' \in V_G} (score(G'))$
    - Si  $score(G_{new}) \geq Score_{max}$  alors
      - $Score_{max} \leftarrow score(G_{new})$
      - $G \leftarrow G_{new}$
    - sinon
      - Continuer Faux
- Tant Que Continuer

**Notations :****Score() :***fonction de score sur les structures possibles* **$V_G$  :***ensemble des GOSCs voisins de graphe G courant***G :***structure finale obtenue par l'algorithme*

**Algorithme K2**

Pour  $i = 1$  à  $n$

$pa_i \leftarrow \emptyset$

$sc_{old} \leftarrow sc(i, pa_i)$

$OK \leftarrow vrai$

Répéter

Chercher  $X_j \in Pred(X_i), pa_i$  qui maximise  $sc(i, pa_i \cup X_j)$

$sc_{new} \leftarrow sc(i, pa_i \cup X_j)$

Si  $sc_{new} > sc_{old}$  alors

$sc_{old} \leftarrow sc_{new}$

$pa_i \leftarrow pa_i \cup X_j$

sinon  $OK \leftarrow faux$

Tant Que  $OK$  et  $|pa_i| < u$

Notations :

**Pred()** : relation d'ordre sur les nœuds  $X_i$

**u** : borne sup. de nombre de parents possibles pour un nœud

$pa_i$  : ensemble des parents du nœud  $X_i$

**sc(i, pa<sub>i</sub>)** : score local

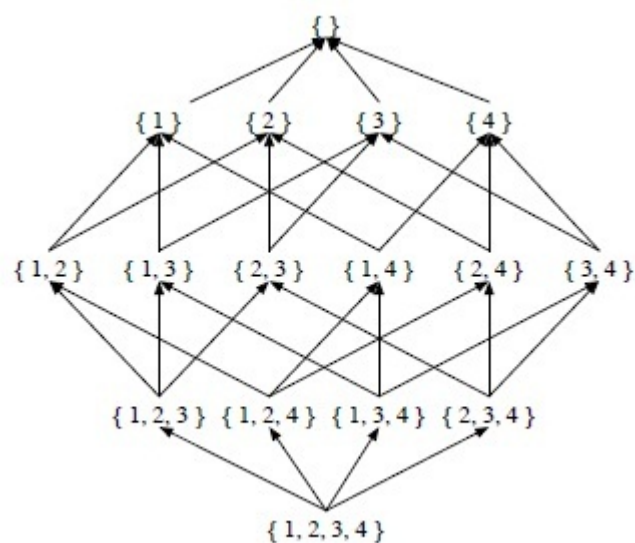


FIGURE 2.2 – Ensemble des sous-reseaux d'un réseau bayésien à quatre nœuds

---

**Algorithme PrgDyn** // Stocke le score du meilleur réseau sur  $S$  dans  $Cache(S)$

---

```

MeilScore  $\leftarrow$  -1
if  $S = \emptyset$  ; then return 0.0
  foreach  $X \in S$  do
    if  $Cache(S - X)$  then
       $s \leftarrow Cache(S - X)$  // récupération du score du meilleur réseau de  $S - X$  s'il existe dans
      le cache
    else
       $s \leftarrow PrgDyn(S - X)$  //si non il calcule le score du meilleur réseau sur  $S - X$ 
       $s \leftarrow s + MeilScore(S, X)$  //ajouter le score du meilleur sous-ens de parents pour  $X$ 
    if  $MeilScore < s$  then
       $Feuille(S) \leftarrow X$ 
       $MeilScore \leftarrow s$ 
   $Cache(S) \leftarrow MeilScore$ 
return  $MeilScore$ 

```

---

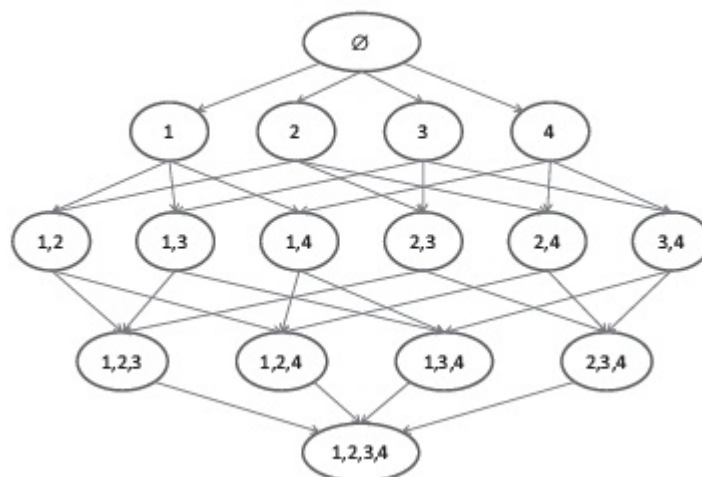


FIGURE 2.3 – Graphe d'ordre pour quatre variables

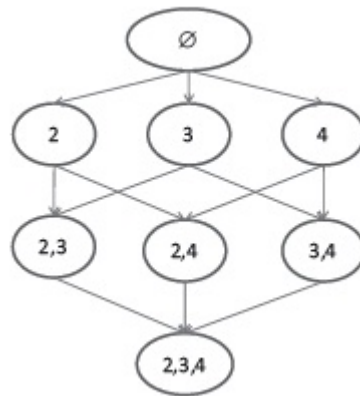


FIGURE 2.4 – Parents possibles de la variable1

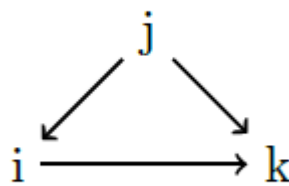


FIGURE 2.5 – Graphe avec 3 nœuds

$i \leftarrow \{\}$	$i \leftarrow \{j\}$	$i \leftarrow \{k\}$	$i \leftarrow \{j, k\}$
0	1	0	0
$j \leftarrow \{\}$	$j \leftarrow \{i\}$	$j \leftarrow \{k\}$	$j \leftarrow \{i, k\}$
1	0	0	0
$k \leftarrow \{\}$	$k \leftarrow \{i\}$	$k \leftarrow \{j\}$	$k \leftarrow \{i, j\}$
0	0	0	1

TABLE 2.1 – Codage des variables de famille d'un graphe avec 3 nœuds

# Chapitre 3

## Application des Réseaux Bayésiens dans le Domaine Médical

### 1 Introduction

Dans le domaine médical, la décision est vue comme étant le centre de l'acte médical. Le processus de cette décision consiste notamment à poser un diagnostic, à proposer un traitement ou à le différer. Ainsi, de nombreuses applications d'aide à la décision ont été développées dans ce domaine. Ces applications sont destinées à aider le personnel de santé dans sa prise de décision. Nous nous intéressons à l'implication des réseaux bayésiens dans le processus décisionnel. Ces techniques consistent à la construction des modèles à partir des informations requises, pour aider le médecin à prendre des meilleures décisions. Dans ce chapitre, nous allons décrire quelques travaux proposés dans ce domaine :

### 2 Etat de l'art

En 2004, un modèle bayésien a été proposé, pour estimer les principales caractéristiques de la transmission de la grippe dans les ménages [47]. Il détaille les risques d'infections dans le ménage et dans la communauté à l'échelle individuelle. L'hétérogénéité des sujets est étudiée en tenant compte à la fois de la susceptibilité individuelle et de l'infectiosité. Le modèle a été appliqué à un en-

semble de données comprenant le suivi des symptômes grippaux. L'estimation des paramètres du modèle de transmission était difficile, car une grande partie du processus infectieux n'a pas été observée : seules les dates auxquelles de nouveaux cas ont été détectés. Pour chaque cas, les données ont été complétées par les dates non observées du début et par la fin de la période infectieuse. Le modèle de transmission était inclus dans une structure hiérarchique à 3 niveaux : (i) le niveau d'observation garantissait la cohérence des données augmentées avec les données observées, (ii) le niveau de transmission décrivait le processus épidémique sous-jacent, (iii) le niveau antérieur spécifié la distribution des paramètres. Dans une perspective bayésienne, la distribution conjointe postérieure des paramètres du modèle et des données augmentées a été explorée par échantillonnage des chaînes de Markov.

Le concept de syndrome en médecine traditionnelle chinoise (MTC) est un énorme système non linéaire, ouvert et compliqué. La différenciation des syndromes en MTC appartient à la science cognitive et noétique. Pour mettre en place un nouveau système de différenciation du syndrome basé sur les éléments clés du syndrome, il est nécessaire que les praticiens de la MTC favorisent la capacité de différenciation et parviennent à un consensus sur la méthode de différenciation. Grâce à la combinaison de l'expérience et des modèles de calcul, en 2006, le réseau bayésien a été utilisé pour étudier la relation entre les éléments clés du syndrome et les symptômes [49], ainsi que la relation entre différents éléments clés, dans laquelle le résultat du diagnostic informatique était identique à celui d'un médecin expérimenté en médecine traditionnelle chinoise. L'étude a montré que le réseau bayésien est une méthode efficace pour traiter les informations sur les symptômes et les signes de différenciation du syndrome.

En 2007, les auteurs proposent un nouvel algorithme de sélection d'attribut pour améliorer la précision de la classification de *Naive Bayes*, appliqué à des bases de données médicales [56]. Les résultats des expériences faites sur 17 bases de données médicales suggèrent qu'en moyenne, le nouvel algorithme CHI-WSS a donné les meilleurs résultats. L'algorithme proposé utilise la discrétisation et simplifie la sélection des attributs basée sur l'approche *wrapper*, il réduit la dimensionnalité des attributs par l'élimination des non pertinents et des moins per-

tinents à l'aide de statistiques du *Khi – deux*. Dans la phase d'expérimentation, La précision de la classification et la courbe de ROC ont été utilisées comme critères d'évaluation.

Un modèle d'essai de randomisation a été proposé en 2008, ce modèle adaptative est basé sur les résultats collectés à partir des patients atteints d'un cancer du poumon au stade avancé [46]. Des échantillons de biopsie de base ont été prélevés sur tous les patients pour une évaluation des biomarqueurs avant la randomisation des traitements. Le critère principal d'évaluation de cette étude est le taux de contrôle de la maladie dans huit semaines après la randomisation. Le modèle bayésien est utilisé pour caractériser ce taux. Les patients sont randomisés de manière adaptative dans l'un des quatre traitements, avec un taux de randomisation basé sur le taux de contrôle de la maladie, ce taux est actualisé à partir des données accumulées dans l'essai. Pour chaque profil de biomarqueurs, les traitements hautement performants présentent des taux de randomisation plus élevés, et inversement. Une règle d'arrêt est mise en œuvre pour suspendre la randomisation des traitements peu performants.

En 2008, un classifieur bayésien a été développé pour classer les patients en fonction de ses caractéristiques statistiques extraites à partir des signaux ECG [51]. Les variables de ces caractéristiques sont converties en une formes symboliques discrètes par seuillage, afin de réduire la dimensionnalité du signal. Cela simplifie le calcul des tables de probabilités conditionnelles pour le classificateur et les rend plus petites. Deux méthodes de construction de réseau à partir de données ont été développées et comparées : la première utilise une recherche gloutonne de *hillclimb*, et la seconde utilise un algorithme génétique (GA).

Une nouvelle approche de la déconvolution aveugle d'images 2D ultrasonores dans un cadre bayésien a été présentée [52]. La déconvolution homomorphique est parmi les méthodes les plus efficaces dans ce domaine, qui repose sur l'hypothèse selon laquelle la fonction d'étalement ponctuel (FEP) et le signal tissulaire se situent dans différentes bandes du domaine du cepstre, ce qui n'est pas tout à fait vrai. De plus, le dépliage de phase 2D est nécessaire dans la cartographie homomorphique 2D, un problème mal posé et sensible au bruit. Ici, les deux limi-

tations sont évitées en utilisant une déconvolution itérative aveugle, à savoir l'algorithme de *VanCittert* avec refonte. La déconvolution homomorphique simplifiée est utilisée uniquement pour l'estimation initiale. L'algorithme est appliqué à l'ensemble de l'image radiofréquence. Des tests sur des images synthétiques et cliniques ont montré que cette méthode donne des résultats stables avec une résolution spatiale clairement supérieure et des structures tissulaires mieux définies.

En 2009, un autre travail a introduit la notion des requêtes probabilistes généralisées pour les Réseaux Bayésiens Dynamiques (RBD), et présente aussi deux algorithmes permettant de répondre à ces requêtes [48]. Ces requêtes généralisées offrent la possibilité de raisonner sur des ensembles de trajectoires via le RBD. Ces algorithmes peuvent également être considérés comme des classifieurs bayésiens pour la vérification du modèle statistique. Après avoir effectué des expérimentations sur deux tâches de décision médicale, ils ont constaté que le modèle produit des bons résultats.

Durant cette année, un autre travail porte sur le développement d'algorithmes de calcul destinés à aider les médecins à prendre des décisions [60]. Où les auteurs proposent un modèle de diagnostic médical fondé sur le principe bayésien naïf flou (BNF), qui étend l'approche bayésienne floue proposée par *Okuda* [96]. L'application du BNF nécessite un entretien avec un médecin pour définir un système d'information orthogonal des symptômes flous. Aux fins d'élaboration et d'élicitation des caractéristiques, l'algorithme est appliqué à un jeu de données simulé simple et comparé à l'approche conventionnelle naïve de Bayes (NB). En tant qu'évaluation préliminaire de la BNF dans un scénario réel, la comparaison est répétée sur un ensemble de données floues de 81 patients diagnostiqués avec des maladies infectieuses. L'étude de cas sur l'ensemble de données simulées montre que BNF peut être optimal pour le diagnostic de patients avec des informations imprécises et floues, en raison des caractéristiques suivantes : 1) il peut modéliser les informations selon lesquelles les valeurs de certains attributs sont sémantiquement plus proches d'autres attributs, et 2) il offre un mécanisme pour tempérer les exagérations dans les informations des patients. Bien que l'algorithme nécessite des données d'apprentissage précises, son utilité pour les données d'apprentis-

sage floues est justifiée. Ceci est confirmé par l'étude de cas sur l'ensemble de données des maladies infectieuses, qui indique l'optimalité de BNF par rapport au NB.

Durant la même année [62], un algorithme de sélection hybride des attributs (CHI-WSS) a été proposé, il permet de réduire la dimensionnalité en supprimant les données non pertinentes. Ce procédé augmentant la précision de l'apprentissage et en améliore la compréhensibilité des résultats. Les expérimentations suggèrent qu'en moyenne, cet algorithme a donné les meilleurs résultats par rapport aux techniques individuelles avec filtre, ainsi qu'aux méthodes de sélection de caractéristiques basées sur le wrapper. L'algorithme proposé est un processus en plusieurs étapes ; au début il discrétise des données continues, puis élimine les caractéristiques les moins importantes et les moins pertinentes et utilise enfin un algorithme glouton pour la sélection des attributs. L'évaluation empirique est faite sur des bases de données disponibles publiquement, où deux mesures pour comparer les performances de ce modèle ont été utilisées, à savoir : l'exactitude de la classification (ou taux d'erreur) et la courbe de ROC. Les expérimentations démontrent que le proposé qui utilise un classifieur naïf bayes, en moyenne est plus efficace que les autres modèles, telles que la régression logistique et la machine à vecteurs de support.

En 2010, une nouvelle méthode a été proposée, pour l'évaluation de la performance des hôpitaux, des médecins et des autres praticiens de la médecine, dans le but d'améliorer la qualité des soins médicaux. Pour cela, une méthode bayésienne a été utilisée [57]. Les auteurs ont comparé les approches fréquentiste et bayésienne à l'aide de données utilisées par le ministère de la Santé de l'État de *New York*. Ces données sont collectées à partir des rapports annuels, selon lesquels les hôpitaux de profil permettent de pratiquer un pontage coronarien. Les comparaisons montrent que la méthode bayésienne produit de meilleurs résultats en comparaison avec d'autre modèle.

Un autre travail proposé en 2010, il explique le rôle du théorème de Bayes et les réseaux bayésiens dans une affaire de négligence médicale introduite par un patient victime d'un accident vasculaire cérébral à la suite d'un test diagnostique

invasif [58]. L'allégation de négligence reposait sur le principe qu'un test alternatif (non invasif) aurait dû être utilisé, car il comportait un risque moins élevé. L'affaire soulève un certain nombre de préoccupations générales et largement applicables concernant le processus de prise de décision au sein de la profession médicale, notamment l'éthique du consentement éclairé, les obligations en matière de soins aux patients lorsque des erreurs sont commises et le problème de recherche consistant à se concentrer sur les «vrais positifs» tout en ignorant les «faux positifs». L'une des préoccupations immédiates est de savoir comment présenter au mieux les arguments bayésiens de manière à ce qu'ils puissent être compris par des personnes qui hésiteraient normalement devant des équations mathématiques. Ils ont présenté des modèles bayésien purement visuelles qui ne nécessitent aucune connaissance mathématique. L'approche prend en charge un grand nombre de scénarios alternatifs, rend toutes les hypothèses faciles à comprendre et offre des avantages potentiels significatifs pour de nombreux domaines de la prise de décision médicale.

Ce nouveau travail propose un critère alternatif dérivé de l'erreur des risques de classification bayésienne pour la segmentation d'images [59]. Le modèle proposé introduit une force basée sur la région déterminée par la différence des densités d'images postérieures pour les différentes classes. Un terme basé sur la probabilité a priori dérivée du nombre d'information de *Kullback-Leibler* et un terme de régularité ont été adoptés pour éviter la génération de signaux excessivement irréguliers et des petites régions segmentées. En Comparaison à d'autres méthodes, l'approche proposée produit une classification optimale des pixels et des estimations de probabilités antérieures ; ainsi, l'approche est plus fiable en théorie et en pratique. Les expériences montrent qu'il est capable d'extraire les formes compliquées et montrent aussi sa robustesse pour différents types d'images médicales. De plus, l'algorithme peut être facilement extensible pour une segmentation multiphase.

L'objectif de ce travail est d'explorer la mise en œuvre d'un modèle bayésien pour la détection du cancer du sein [64]. Il rassemble des entités importantes tels que les radiologistes, les scientifiques en traitement d'images, les spécialistes en bases de données et les mathématiciens appliqués sur une plate-forme commune.

Il est expliqué qu'en exploitant les indépendances conditionnelles entraînées par des chaînes d'influence, il est possible de représenter un grand exemple dans un réseau bayésien en utilisant peu d'espace, et il est souvent possible d'effectuer une inférence probabiliste dans une durée acceptable. Les étapes pour la réalisation d'une mise en œuvre du modèle bayésien pour la détection assistée par ordinateur en mammographie, sont décrites dans ce travail. Les chercheurs ont l'intention de concevoir une interface entre l'algorithme d'apprentissage en réseau bayésien et les radiologues, afin que les radiologistes puissent avoir une interaction avec le système en n'étiquetant qu'un petit nombre d'images informatives présentées par l'algorithme d'apprentissage actif.

Un système est utilisé pour le diagnostic médical, basé sur l'ontologie et le réseau bayésien [50]. Il a deux objectifs : (1) la séparation de la connaissance du domaine des informations probabilistes et (2) la création d'une interface utilisateur intuitive. L'architecture de la structure comporte trois couches : connaissances, modèle d'incertitude et interface utilisateur. Les experts du domaine ont contribué à la construction de l'ontologie, cette dernière contient que les concepts et les relations axées uniquement sur la connaissance du domaine. Le modèle d'incertitude est le réseau bayésien et les probabilités des états de variables sont stockées dans un référentiel de profils. Le médecin utilise l'interface alimentée avec l'ontologie du domaine et le modèle d'incertitude. L'application a été testée sur un modèle d'échantillon de médicament pour le diagnostic de la maladie cardiaque.

En 2011, L'apprentissage causal hybride des RB est appliqué dans le domaine médical [53], les auteurs ont présenté des méthodes pour intégrer les connaissances d'experts dans le processus d'ASBR, ils ont examiné un certain nombre d'approches pour combiner ces structures lors de la phase d'apprentissage, et présenter de nouvelles techniques pour évaluer les résultats. Le workflow détaillé d'ingénierie des connaissances s'avère utile pour la structuration d'un processus de développement de RB itératif complexe. Après des études faites sur des cas médicaux issus des jeux de données IOWA,

les matrices de contiguïté indiquent clairement que l'approche du tri partiel des variables en niveaux est la plus efficace, il facilite la découverte du modèle.

Dans ce travail, les auteurs ont proposé un modèle bayésien, afin d'automatiser les tâches exigées par le Bureau de la gestion du personnel des États-Unis [63]. Il a été demandé à tous les régimes de soins de santé participant au Programme de prestations de santé et des employés fédéraux, d'offrir des avantages pour la santé mentale et la toxicomanie équivalents aux avantages médicaux généraux. L'évaluation initiale au cours de la période d'observation de quatre ans, a révélé qu'en moyenne, la parité ne s'est traduite par aucune augmentation des dépenses, ni par une utilisation accrue des services. Cependant, certains groupes d'inscrits ont peut-être bénéficié de la parité plus que d'autres.

Pour remédier à cela, ce travail propose un modèle bayésien à deux niveaux de classes latentes pour caractériser l'effet de la parité et les dépenses dans la santé mentale.

Dans chaque classe, les auteurs ont adapté un modèle à effets aléatoires en deux parties, pour modéliser séparément la probabilité d'utilisation de la santé mentale ou de la toxicomanie et les trajectoires de dépenses moyennes utilisées par les services. L'analyse a identifié trois catégories de sujets : un groupe de personnes dépensant peu qui avaient tendance à être des hommes, utilisaient relativement peu de services et diminuaient avec le temps ; un groupe de consommateurs modérés, principalement des femmes, dont l'utilisation et les dépenses moyennes ont augmenté à la suite de l'introduction de la parité ; et un groupe de grands consommateurs qui avaient tendance à avoir une utilisation chronique des services et des habitudes de dépenses constantes. Après avoir examiné les régions de densité de probabilité les plus élevées associées à 95% des changements attendus dans l'utilisation et les dépenses pour chaque classe, ils ont confirmé que la parité n'avait un impact que sur la classe des dépenses moyennes.

Un autre système utilise une ontologie en tant que référentiel des connaissances médicales, est construit pour le diagnostic médical [68]. Cette ontologie peut être élaborée par le médecin spécialiste au cours de l'utilisation du système. À cette fin, l'utilisateur utilise une interface graphique par laquelle les nouvelles pathologies et symptômes possibles sont introduits, ainsi que les probabilités qui caractérisent leurs dépendances. Un expert fournit un support pour valider la cohérence des données introduites, ainsi que pour déduire les propriétés impli-

cites et les relations entre elles, ce qui offre un moyen de développer davantage les connaissances. L'ontologie est traitée de la manière la plus conventionnelle, en utilisant les standards OWL. Afin d'effectuer l'analyse bayésienne, des corrélations probabilistes entre les observations et les pathologies ont été introduites. Ils ont utilisé un réseau bayésien à deux niveaux et une structure prédéfinie pour les réseaux de décision. Pour générer et évaluer le modèle décisionnel, ils ont utilisé une bibliothèque standard, dont les fonctionnalités sont appelées en fonction du type de réseau que le médecin peut utiliser. Ceci est spécifié via une interface système conviviale qui assiste l'opérateur dans la formulation du diagnostic et donc du réseau impliqué. Un prototype du système est actuellement utilisé par les spécialistes qui ont contribué à la définition des exigences du système, qui ont aussi un impact important sur la construction de la structure du RB.

Un modèle est utilisé pour la détection du cancer du sein [91], qui est basé sur les lésions détectées par la mammographie, la base de données Wisconsin Diagnostic Breast Cancer (WDBC) a été utilisée pour la construction du modèle bayésien. Cette base de données contient un grand nombre de descripteurs (31 variables), Le modèle construit à partir de cette base de données a offert des résultats satisfaisants, démontrant l'efficacité de l'approche probabiliste dans le domaine médical.

Une technique de modélisation statistique a été proposée en 2012, appelée modèle HARM (Hierarchical Association Rule Model) [55], ce modèle prédit les conditions médicales futures possibles du patient, en fonction de ses antécédents. La technique repose essentiellement sur un modèle bayésien hiérarchique permettant de sélectionner des règles d'association prédictives (telles que : condition 1 et condition 2  $\rightarrow$  condition 3) à partir d'un vaste ensemble de règles candidates. Cette méthode permet de fournir des prévisions spécifiques à un patient donné, même lorsque peu d'informations sur les antécédents médicaux du patient sont disponibles, car elle utilise les conditions de nombreux patients similaires.

Dans cet article, les auteurs ont examiné la sensibilité à la précision des paramètres des modèles bayésien [54]. Les modèles de tests incluent *HeparII*, qui est un réseau bayésien dédié pour le diagnostic des troubles du foie, ainsi six autres

réseaux de diagnostic médical construits à partir de jeux de données médicales disponibles dans le référentiel *Irvine Machine Learning*. Ils ont supposé que les paramètres du modèle d'origine sont parfaitement précis, ils ont réduit systématiquement leur précision en les arrondissant progressivement à des échelles plus plates et en vérifiant l'incidence de cet arrondi sur la précision du modèle. Les résultats des expériences démontrent que, tant que les zéros sont évités parmi les paramètres du modèle, la précision du diagnostic des modèles de réseau bayésien ne souffre d'aucune diminution de précision de leur paramètres.

Les dépenses médicales en fin de vie représentent une proportion importante de toutes les dépenses de santé. Ces coûts ont été étudiés à l'aide des coûts des services de réclamations d'assurance-maladie et des causes de décès (CoD) issues des certificats de décès [66]. En l'absence d'identifiant unique reliant les deux jeux de données, les variables communes ont identifié des correspondances uniques pour seulement 33% des décès. Ils ont échantillonné la distribution postérieure commune des paramètres du modèle et les permutations qui relient les observations des deux fichiers de chaque cellule. Les modèles de liaison comprenaient la régression de la localisation du décès sur CoD et d'autres paramètres, ainsi que la régression des mesures de coûts avec un modèle de données manquant monotone sur la CoD et d'autres caractéristiques démographiques. Les permutations ont été échantillonnées en énumérant la distribution exacte pour les petites cellules et par l'algorithme de *Metropolis* pour les grandes cellules. Les structures de données matricielles fragmentées ont permis des calculs efficaces, malgré la taille de la base de données (environ 1,7 millions d'exemples). La procédure génère  $m$  jeux de données dans lesquels les correspondances entre les deux fichiers sont fournies. Les  $m$  ensembles de données peuvent être analysés indépendamment et les résultats peuvent être combinés à l'aide des règles d'imputation multiple.

Un modèle probabiliste a été proposé en 2014 [92], il est dédié pour la reconnaissance d'un type d'arythmies cardiaques qui est le battement ventriculaire prématuré, ces maladies cardio-vasculaires présentent un taux de décès élevé dans le monde, d'où la nécessité d'une détection pour minimiser leur risque. La base de données utilisée pour l'apprentissage est La base MIT-BIH, Les résultats obtenus étaient satisfaisants.

Il faut noter qu'avant la phase d'apprentissage, une méthode de type wrapper a été utilisée, pour extraire que les variables les plus pertinentes, ce qui facilite les tâches de constructions de ce modèle.

Un système permettant d'obtenir des modèles de données à l'aide du théorème naïf bayésien a été proposé dans [65]. Les expérimentations sont faites sur des données recueillies auprès de centres de soins. Ce système révèle les risques des différentes maladies dermatologiques et détermine également le pourcentage d'occurrence de chaque maladie.

Une nouvelle technique d'entraînement regroupant l'apprentissage avec correction d'erreur a été proposé en 2014 [67], où la distribution de probabilité postérieure des pondérations et la corrélation de rang de *Goodman-Kruskal* ont été assemblées dans une stratégie d'apprentissage bayésienne. Cette étude avait deux objectifs principaux : premièrement, développer une nouvelle technique d'apprentissage basée à la fois sur le paradigme bayésien et la propagation d'erreurs, et deuxièmement, évaluer son efficacité. Les performances du modèle proposé sont comparées à celles obtenues par des algorithmes traditionnels d'apprentissage automatique, utilisant des bases de données médicales réelles sur le cancer du sein et du poumon, le diabète et les crises cardiaques. Dans l'ensemble, les résultats indiquent que la nouvelle méthode d'apprentissage surpasse les techniques classiques presque à tout l'égard.

En 2016, une méthode rigoureuse et reproductible pour la construction de modèles de réseaux bayésiens efficaces a été proposé pour l'aide à la décision médicale [61]. Ce modèle est construit à partir de questionnaires et d'interviews complexes, non structurés et incomplets collectés des patients, contenant inévitablement des exemples de réponses répétitives, redondantes et contradictoires. Il exploite aussi les connaissances des spécialistes dans le développement d'un réseau bayésien. Il s'est confirmé que ce modèle bayésien peut être utilisé pour l'analyse des interventions; les auteurs ont démontré aussi que l'utilisation de données seules pour connaître la structure et les paramètres du modèle est souvent insatisfaisante, même lorsque de nombreuses données sont disponibles.

Dans le travail [106], les auteurs ont étudié L'engagement des patients pour améliorer la qualité des soins de santé. Cependant, les relations globales entre les mesures de participation des patients et les résultats de la santé ne sont pas claires. Cette étude a pour objectif d'intégrer les connaissances d'experts et les données d'enquête afin d'identifier les mesures ayant des associations étendues avec d'autres variables et pouvant être hiérarchisées pour engager les patients. il ont utilisé les données issus de l'Enquête internationale sur la santé (IHPS) de 2014, qui fournissait des informations sur les personnes âgées de 11 pays et fournissait des informations détaillées sur les caractéristiques des patients. L'engagement ou le soutien des patients a été mesuré à l'aide de huit variables, notamment les choix de traitement des patients, leur implication et l'établissement des priorités de traitement. Trois types de soins ont été identifiés : soins primaires, soins spécialisés et maladies chroniques. Les experts étaient des médecins spécialisés dans un domaine des soins. Les maladies chroniques comprenaient huit maladies chroniques étudiées. Des connaissances spécialisées ont été utilisées pour faciliter la sélection des variables. Ils ont utilisé des modèles de réseau bayésiens pour représenter les connaissances. D'après les expérimentations, les stratégies d'implication des patients peuvent dépendre des types de soins et des soins de maladies chroniques.

### 3 Conclusion

Les réseaux bayésiens sont parmi les méthodes les plus populaires utilisées pour l'aide au diagnostic médical, dans ce chapitre nous avons cité quelques applications qui utilisent ce modèle dans le domaine médical, ainsi que leurs efficacité en comparaison avec d'autres méthodes dans la littérature.

Dans le chapitre suivant nous montrons nos contributions pour l'ASRB, qui peut être considéré comme étant l'étape la plus complexe dans le processus de construction d'un RB, vu le nombre exponentiel des GOSCs possibles relatifs à un ensemble de variables. Et nous montrons aussi l'efficacité de ces améliorations apportées à l'algorithme *B&B* standard utilisé pour l'ASRB.

# Chapitre 4

## Approches Proposées

### 1 Introduction

Ce dernier chapitre est destiné essentiellement à nos principales contributions ; nous commençons par une description détaillée de l'algorithme *B&B* standard, dédié pour l'ASRB, ensuite nous montrons les limitations de ce dernier. Puis nous expliquons nos améliorations que nous avons apporté à cet algorithme avec un exemple illustratif. Les bases d'exemples utilisées dans les expérimentations sont extraites à partir du site UCI [39]. Parmi ces bases de données, trois maladies avec ses bases d'exemples sont décrites en détails dans l'annexe A.

Nos améliorations ont été justifiées par des comparaisons faites avec la méthode standard ainsi qu'avec d'autres méthodes décrites précédemment dans le deuxième chapitre.

### 2 Problématique

Dans cette section, nous allons décrire l'algorithme proposé par De Campos et al., [37] pour l'ASRB, ils ont appliqué la méthode B&B pour l'extraction de la meilleure structure d'un RB. L'avantage principale de cette méthode en comparaison avec d'autre méthode exacte est sa capacité de produire des solutions complètes approximatives avant la fin de l'exécution, et par conséquent elle peut être arrêtée à tout moment, et fournir une solution approximative (la meilleure solution réalisable trouvée jusqu'à présent). Durant la description de cet algorithme,

nous utilisons un exemple pour une meilleure explication.

---

**Algorithm Std-B&B** : Pseudocode of the standard B&B algorithm for ASRB

---

```

1 :  $S_{best} \leftarrow -\infty$ 
2 :  $Insert((G, H, S), Q)$ 
3 : While NotEmpty(Q) do
4 :  $(G_{cur}, H_{cur}, S_{cur}) \leftarrow take(Q)$ 
5 : If  $(S_{cur} > S_{best})$  then
6 :   If  $(DAG(G_{cur}))$  then
7 :      $S_{best} \leftarrow S_{cur}$ 
8 :      $G_{best} \leftarrow G_{cur}$ 
9 :   Else
10 :     $cyc \leftarrow FirstCycle(G_{cur})$ 
11 :    For all  $arc(u \rightarrow v) \in cyc$  do
12 :       $Prohibited((u \rightarrow v), H_{cur})$ 
13 :       $(G, S) \leftarrow GetRestGraph(G_{cur}, H_{cur}, Cache(v, PA_v))$ 
14 :       $Insert((G, H_{cur}, S), Q)$ 
15 :       $MustExist((u \rightarrow v), H_{cur})$ 
16 :       $Prohibited((v \rightarrow u), H_{cur})$ 
17 :    Endfor
18 :  Endif
19 :
20 : Endif
21 : Enddo

```

---

Nous commençons par une description des structures de données utilisées dans cet algorithme (Algorithme **Std-B&B**) :

1. **Le cache :**

Le cache contient tous les scores des parents possibles pour chaque variable, il permet d'éviter le recalcul du même score local plusieurs fois.

Soit  $n$  le nombre de variables, et  $2^{n-1}$  le nombre des parents possible pour chaque variable, alors, la taille du cache est  $n \times 2^{n-1}$ , l'espace de recherche est égale  $2^{n-1} \times \dots \times 2^{n-1} = 2^{n \times (n-1)}$ .

Les propriétés proposées par decampos et al. [37] ont pour but de réduire l'espace de recherche(voir section 1.8).

Soit l'exemple défini dans le tableau suivant(Tableau 4.1), contenant le cache calculé à partir d'une base de données de 5 variables.

Pour simplifier l'exemple, on suppose qu'on a utilisé les propriétés vu précédemment(Section 1.8), donc la taille du cache devient 14 au lieu de  $80(5 \times 2^4)$ .

TABLE 4.1 – Cache des Scores locaux pour chaque variable

Nœud	Parents	Score
1	{ 3,4 }	5
1	{ 3 }	3
1	{ 4 }	2
1	{ }	1
2	{ 1 }	2
2	{ }	1
3	{ 2 }	3
3	{ }	2
4	{ 1,5 }	7
4	{ 5 }	6
4	{ 1 }	4
4	{ }	3
5	{ 4 }	8
5	{ }	7

Le tableau4.1 contient tous les parents possibles avec leurs scores, pour toutes les variables, et pour chaque variable, les lignes sont triées par ordre décroissant suivant le score local.

## 2. Le graphe initial G :

Soit  $G = (V, E)$  le graphe initial, où  $|V| = n$  (voir exemple figure4.1 ). Ce graph est crée à partir de la meilleure configuration de toutes les variables, qui est généralement un graphe cyclique.

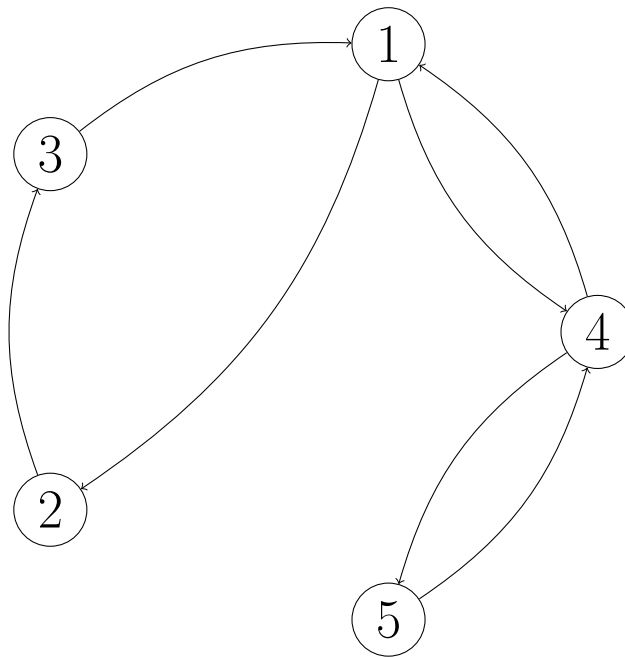


FIGURE 4.1 – Graphe initial G.

$G$  est représenté par une matrice  $A$  ( $n \times n$ ) (Tableau 4.2) booléenne avec :  $A_{ij} = 1$  s’il existe un arc entre le nœud  $i$  et  $j$  et 0 sinon.

TABLE 4.2 – Matrice A

	1	2	3	4	5
1	0	0	1	1	0
2	1	0	0	0	0
3	0	1	0	0	0
4	1	0	0	0	0
5	0	0	0	1	0

3. **Le score S :**

$S$  est le score global de  $G$ , qui représente la somme des scores locaux de toutes les variables du problème donné, il est utilisé comme borne inférieure de la solution partielle.

4. **La Matrice de control H :**

$H$  est une matrice ( $n \times n$ ) initialement vide, pour chaque arc entre les nœuds, un jalon de son état est indiqué par la matrice  $H$ , qui peut être soit présent, absent ou libre. Cette matrice est utilisée par la méthode  $B\&B$  pour contrôler la recherche de la meilleure structure d’un  $RB$ , Chaque branche dans

l'arbre de recherche contient une matrice, qui permet de guider la recherche et d'éviter les redondances. Initialement tous les arcs sont libres.

#### 5. La file Q :

Dans le problème de recherche de la meilleure structure d'un réseau bayésien, l'objectif principal de la méthode *B&B* est de maximiser le score global  $S$  (borne inférieure).

La stratégie du meilleur d'abord est appliquée dans le parcours des solutions partielles, elle consiste à explorer des sous-problèmes possédant la meilleure borne, elle permet aussi d'éviter l'exploration des sous-problèmes possédant une mauvaise évaluation par rapport à la valeur optimale. Par conséquent, elle permet d'obtenir la solution recherchée en un temps raisonnable.

Pour l'implémentation de cette stratégie, une file de priorité Q contenant des triplets  $(G,H,S)$  est utilisée, où les triplets sont triés par ordre décroissant selon  $S$ .

#### 6. La meilleure solution réalisable $G_{best}$ :

Durant l'exécution de l'algorithme *B&B*, plusieurs solutions réalisables sont obtenues, le graphe et le score de la meilleure solution est enregistré respectivement dans  $G_{best}$  et  $S_{best}$ . Ce score est considéré comme borne inférieure. Initialement, il prend la valeur  $-\infty$

L'objectif principal de cet algorithme est de maximiser une fonction de score décomposable.

Après la construction du graph initial à partir de la meilleure configuration de chaque variable, l'algorithme *B&B* commence par l'exploration de ce graphe. Dans chaque itération, le triplet courant  $(G_{cur}, H_{cur}, S_{cur})$  est récupéré de la file d'attente Q (voir ligne 4). Ensuite, l'algorithme teste si le score actuel  $S_{cur}$  peut améliorer le meilleur score enregistré dans  $S_{meilleur}$  (voir ligne 5). Si c'est le cas, la ligne 6 vérifie si ce graphe choisi est un *GOSC* (solution réalisable). Et la meilleure solution est mise à jour aux lignes 7,8 si c'est nécessaire. Sinon, le graphe doit avoir des cycles. Le premier cycle du graphe actuel  $G_{cur}$  est candidat pour l'élimination (Ligne 10), en obligeant certains arcs à être absents/présents. Ainsi, pour

chaque arc  $(u \rightarrow v)$  de ce cycle :

- un nouveau sous-graphe  $G$  est formé par les arcs demandés par  $H_{cur}$ , et son score  $S$  est calculé (le cache  $C(v, PA_v)$  est utilisé pour éviter de recalculer les scores), où seul le score local du nœud  $v$  est modifié par le nouveau dans le score global  $S$ , lignes 12,13 ;
- le triplet  $(G, H_{cur}, S)$  est inséré dans  $Q$ , (ligne 14) ;
- l'arc  $(u \rightarrow v)$  est marqué présent dans  $H_{cur}$  ; cette procédure force les branches que nous créons à être disjointes les unes des autres (ligne 15), ce qui évite les graphes redondants ;
- l'arc  $(u \rightarrow v)$  est marqué absent dans  $H_{cur}$  ; cela évite le cycle  $(u \rightarrow v \rightarrow u)$ , (ligne 16) ;

À la ligne 5, si le score du triplet actuel  $(G_{cur}, H_{cur}, S_{cur})$  est inférieur à la meilleure solution trouvée jusqu'à présent (la borne inférieure  $S_{best}$ ), la branche associée est complètement ignorée, puisqu'elle ne peut contenir la solution optimale.

Cet algorithme s'arrête lorsque la file d'attente  $Q$  est vide (ligne 3), et retourne le meilleur GOSC  $G_{best}$  avec un score global  $S_{best}$ .

On remarque que cet algorithme possède trois inconvénients majeurs :

1. La stratégie du parcours utilisé dans cet algorithme est "*le meilleur d'abord*", elle consiste à explorer les sous-graphes ayant le meilleur score (borne supérieure), ce qui permet d'éviter l'exploration des sous-graphes qui ne mènent pas à des solutions optimales, qui ont un mauvais score. Cette stratégie de parcours produit moins de solutions approximatives. Cependant l'avantage principale de cette méthode est sa capacité de produire des solutions approximatives avant la fin de l'exécution, et peut être arrêtée à tout moment, et fournir une solution approximative (la meilleure solution réalisable trouvée jusqu'à présent), par conséquent, la stratégie du parcours de cette méthode doit être améliorée afin de garder cet avantage.
2. La procédure du branchement dans un graphe influe sur la complexité spatiale de cet algorithme, car tous les sous-graphes résultants sont sauvegardés dans la file, ils sont obtenus après la suppression d'un cycle, et le nombre de ces sous-graphes est égal à la taille du cycle choisi pour la sup-

pression. Un graphe peut avoir plusieurs cycles avec des tailles différentes, cet algorithme choisit le premier cycle trouvé afin de le supprimer, alors que le choix du cycle adéquat parmi ces cycles peut réduire considérablement la complexité spatiale.

3. L'algorithme standard B&B utilise le score du graphe courant comme borne supérieure. Toutefois, les performances d'un algorithme B& B dépendent de l'efficacité de l'heuristique utilisée pour le calcul de bornes supérieures, qui permet d'écarter un grand nombre de sous problèmes.

## 3 Contributions

### 3.1 Introduction

Afin d'améliorer la performance de l'algorithme standard *B&B* [37], Nous proposons quatre contributions :

1. **Stratégie du parcours :**

La méthode standard se base sur la stratégie du parcours du meilleur d'abord, qui réduit l'efficacité de cette méthode, car parmi les avantages de cet algorithme est son arrêt à tout moment, et fournir une solution approximative, donc il est intéressant de choisir une stratégie de parcours qui permet de produire des solutions approximatives le plus rapidement possible. Pour remédier à cela, nous proposons une combinaison de deux stratégies ; la "*recherche en profondeur*" et "*la recherche du meilleur d'abord*", dont le but de profiter au maximum des avantages de ces deux stratégies.

Pour cela, nous proposons une stratégie de parcours permettant de choisir le graphe ayant le meilleur score parmi ceux qui ont le moins nombre de cycles.

2. **Strategie du branchement :** Le graphe obtenu par la stratégie du parcours peut contenir plusieurs cycles. La méthode de B&B standard choisit le premier cycle candidat pour la suppression. Tous les sous-graphes déduits par suppression de ce cycle sont enfilé dans la file pour les prochaines explorations.

Nous proposons une stratégie pour réduire la complexité spatiale (taille de la file). Elle permet de choisir le cycle ayant le moins d'arc. Et par conséquent, le nombre de sous-graphes enfilés dans la file est réduit, car ce nombre dépend directement de la taille du cycle supprimé.

3. **Opérateur du bornage** : Les performances d'une méthode B& B dépendent de l'efficacité de l'opérateur du bornage, qui est utilisée pour le calcul de la borne supérieure. Dans l'algorithme standard B&B, le score  $S$  du graphe  $G$  est utilisé comme borne supérieure. Afin d'accélérer l'exploration d'arbres de recherche, nous proposons un opérateur du bornage efficace, qui permet de calculer une borne supérieure pour chaque nœud (sous-problème) dans l'arbre.
4. **Parallèle Branch&Bound** : Ce concept de parallélisme représente le fait d'exécution simultanée de plusieurs processus ; le même code sera implémenté sur plusieurs processeurs. Ça permet d'accélérer le traitement. Dans ce travail nous adoptons le modèle d'exploration d'arborescence parallèle.

## 3.2 Stratégies du parcours et du branchement

Dans cette section, nous montrons l'efficacité des deux premières contributions, les stratégies du parcours et du branchement, nous considérons un exemple illustratif, permettant de montrer l'efficacité de ces stratégies proposées, nous expliquons par cet exemple les détails des améliorations apportées à la méthode B&B standard pour l'ASRB.

### 3.2.1 Exemple illustratif

La figure 4.1 représente le graphe initial utilisé dans cet exemple, les scores sont calculés à partir des scores locaux donnés dans le tableau 4.1.

La taille du cache ou bien le nombre des parents dans un réseau bayésien à  $n$  variables est  $n2^{n-1}$ , qui est égale à 80 ( $5 \times 2^4$ ) pour  $n = 5$ . Dans notre exemple nous avons pris que quelques parents soit 14, pour deux raisons : la première se traduit par notre supposition que les propriétés décrites dans la section 1.8 ont été utilisées, la deuxième est dans le but de simplifier notre exemple, car le nombre d'itérations de la méthode  $B&B$  dépend directement de la taille du cache.

L'algorithme précédent [37] est illustré dans la figure 4.2. Après la construction du graphe initial à partir de la meilleure configuration de chaque variable, Le graphe doit contenir 3 cycles ( $W_1$ ,  $W_2$  et  $W_3$ ), son score égale à 25 ( $5+2+3+7+8$ ) considéré comme "*borne supérieure*", le meilleur score obtenu jusqu'à présent est initialisé à  $-\infty$  qui représente la "*borne inférieure*".

L'algorithme *B&B* commence par la suppression du premier cycle trouvé dans le graphe qui est  $W_1$  (il contient 3 arcs). Par conséquent, 3 de nouveaux sous-graphes sont formés. Chacun d'entre eux est stocké dans la file d'attente selon l'ordre décroissant du score.

Les scores de ces sous-graphes sont recalculés en se basant sur le tableau de cache 4.1 comme suit :

- **Elimination de l'arc  $1 \rightarrow 2$**  : le score global est modifié par la soustraction du score local de la variable 2, et le remplacer par le nouveau score local de la même variable sachant que la variable 1 ne figure pas dans la liste des parents, et donc le nouveau score local sera 1 au lieu de 2. Par conséquent le score global est égale à  $24(25-2+1)$ .
- **Elimination de l'arc  $2 \rightarrow 3$**  : le score global est égale à  $24(25-3+2)$ , car le score local de la variable 3 sera 2 au lieu de 3(la variable 2 ne figure pas dans l'ensemble des parents de la variable 3).
- **Elimination de l'arc  $3 \rightarrow 1$**  : le score global est égale à  $22(25-5+2)$ , car le score local de la variable 1 sera 2 au lieu de 5(la variable 3 ne figure pas dans l'ensemble des parents de la variable 1).

Dans l'itération suivante, un graphe parmi les deux graphes ayant le meilleur score(24), est choisit au hasard, et son cycle  $w_2$  est supprimé.

Cette opération se répète pour tous les graphes de la file d'attente. Les GOSCs sont obtenus après la suppression consécutive de  $w_1$ ,  $w_2$  et  $w_3$  ; dans ce cas, le score du GOSCs résultant est comparé au meilleur obtenu jusqu'à présent. S'il est meilleur, le graphe et son score sont enregistrés comme meilleur résultat jusqu'à présent. Dans cet exemple, le score du premier GOSCs obtenu est 22, ce résultat n'a pas été amélioré jusqu'à la fin de l'exécution.

Nous notons, que dans les itérations 6, 7 et 8, les sous-graphes sont complètement ignorées car leurs scores sont inférieurs ou égaux au meilleur obtenu jusqu'à présent (22). Cet algorithme s'arrête lorsque la file d'attente est vide, et retourne le meilleur graphe obtenu.

Nous constatons également que la procédure d'apprentissage se fait en 8 itérations, la taille maximale de la file d'attente est de 5 et le premier GOSC est obtenu après 4 itérations.

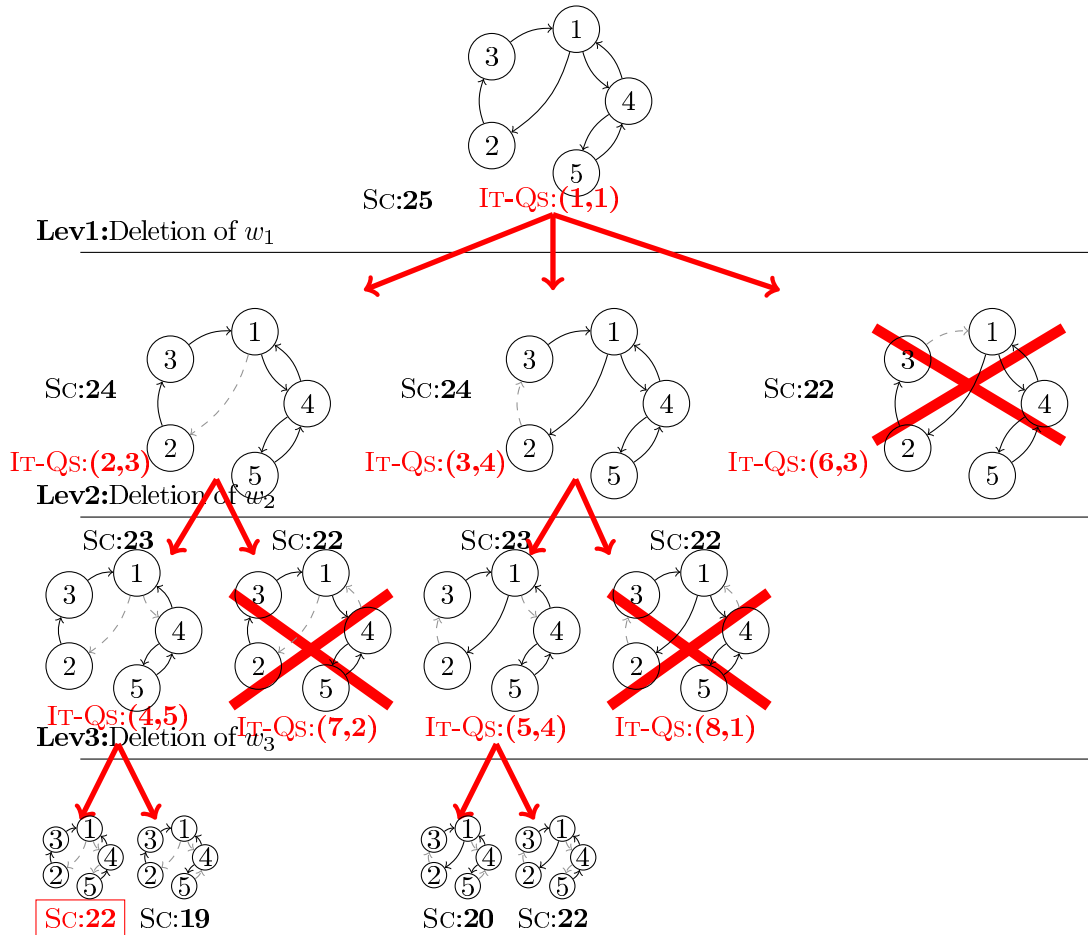


FIGURE 4.2 – Execution de la méthode B&B.

La figure 4.3 illustre la stratégie du parcours proposé (Première contribution). Au cours du processus d'élimination des cycles, il donne la priorité au graphe ayant le meilleur score parmi ceux qui contiennent moins de cycles ;(chaque graphe est stocké dans la file d'abord suivant son nombre de cycles dans l'ordre croissant, puis par le score dans l'ordre décroissant), cette stratégie tente de trouver les GOSCs le plus rapidement possible. Après le choix du graphe pour la prochaine exploration, la procédure élimine le premier cycle trouvé. Nous notons que les mêmes graphes sont obtenus dans les deux stratégies (Figure 4.2, 4.3). Cependant, lors de l'application de cette dernière procédure, le premier GOSC est obtenu après 3 itérations et la taille maximum de la file est de 4.

Cette stratégie accélère la procédure d'apprentissage d'une part, et stocke moins de graphe dans la file d'autre part.

Il faut noter aussi qu'il est très important d'obtenir rapidement les *GOSC*s qui constituent les meilleures solutions actuelles, car l'avantage principale de cette méthode est sa capacité de fournir des solutions approximatives à tout moment, durant l'exécution.

Cette heuristique profite des avantages des deux stratégies de recherche ; la recherche en *profondeur* et la *recherche du meilleur d'abord*. Elle commence par le graphe qui a le meilleur score parmi les graphes ayant moins de cycles, ce qui réduit le nombre d'itérations et la consommation de mémoire pour obtenir des meilleurs résultats.

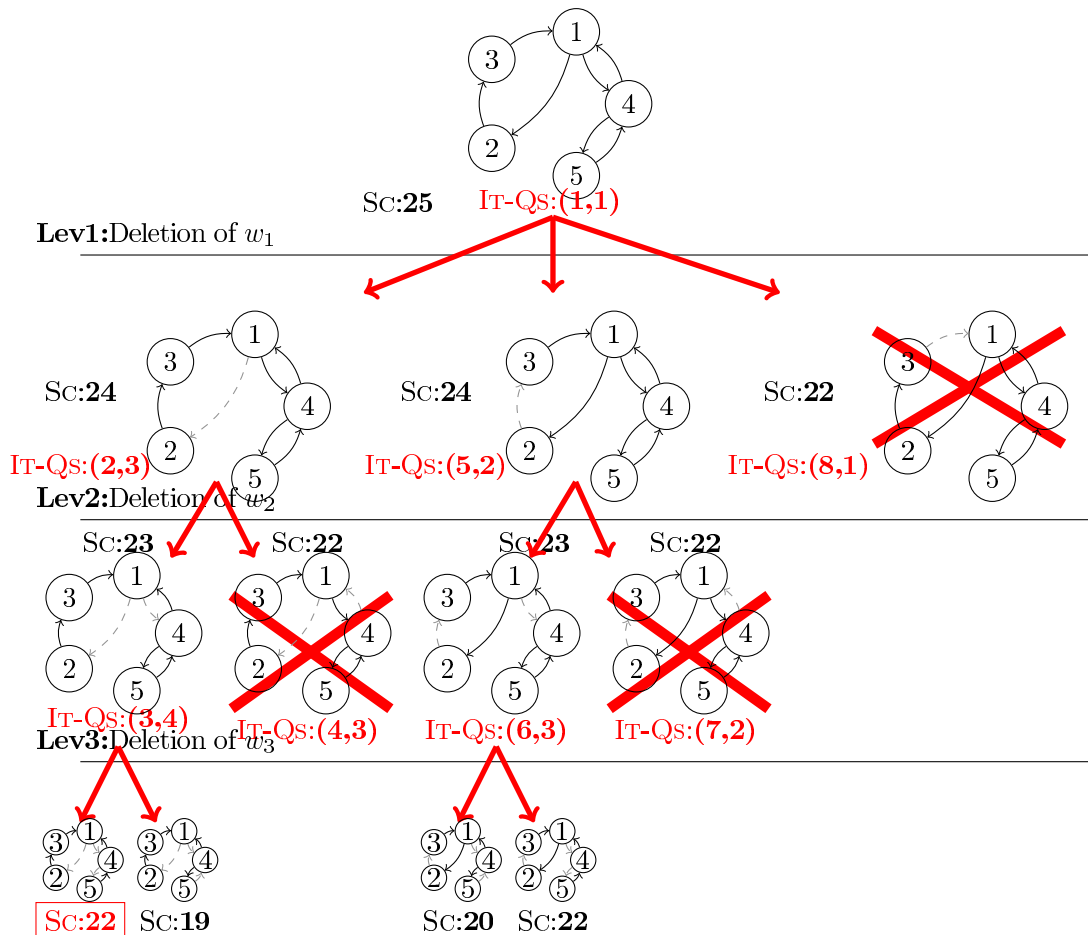


FIGURE 4.3 – Elimination des cycles selon la nouvelle stratégie du parcours.

La figure 4.4 illustre la stratégie du branchement ((Deuxième contribution). Elle sélectionne le graphe présentant le meilleur score, puis élimine le cycle le

plus court de ce graphe. Donc,  $w_2$ ,  $w_3$  et  $w_1$  sont traités consécutivement car  $|w_2| \leq |w_3| < |w_1|$ . Chaque graphe obtenu après l'élimination de ce cycle, est stocké dans la file en fonction de l'ordre décroissant du score. Notez que la taille maximale de la file d'attente est 4. De plus, l'apprentissage de structure est fini dans 7 itérations. Cela est dû au fait que le traitement du cycle le plus long  $w_1$  est laissé à la fin.

L'application de cette stratégie a pour objectif de stocker le moins nombre de sous-graphes dans la file d'attente, c.à.d. à chaque itération elle choisit le cycle le plus court dans le graphe sélectionné, qui sera éliminé par la suite.

Il faut noter que le nombre des sous-graphes résultants après cette élimination est égal au nombre d'arcs contenu dans ce cycle. Par conséquent l'élimination du cycle ayant le moins d'arc réduit la complexité spatiale.

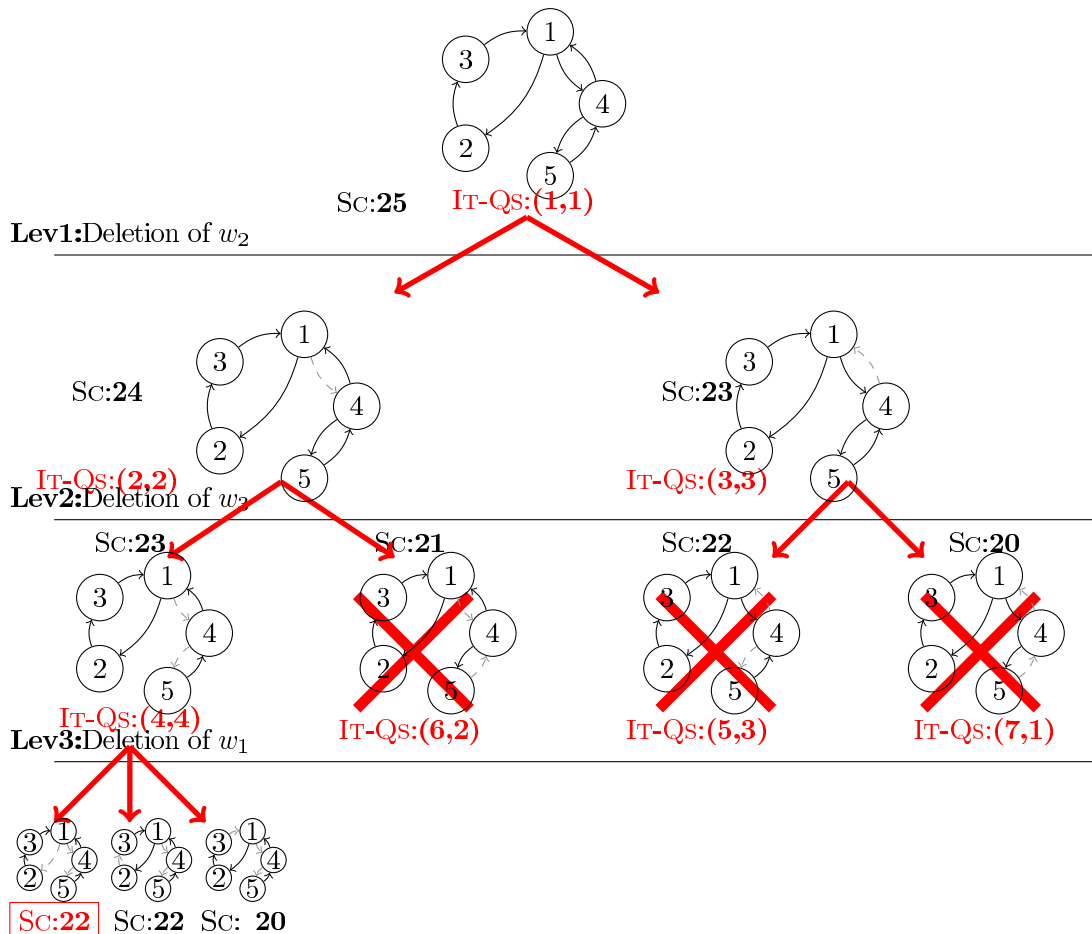


FIGURE 4.4 – Elimination des cycles selon la nouvelle startégie du branchement.

La figure 4.5 peut être considérée comme une combinaison des deux avantages constatés dans les deux contributions précédentes (Figures 4.3, 4.4).

Pour résumer ces deux contributions : la première consiste à choisir le graphe ayant le meilleur score parmi ceux qui contiennent moins de cycles (Figure 4.3) ; la seconde élimine le cycle le plus court du graphe choisi (Figure 4.4).

Nous constatons une amélioration significative, qui se justifie par une réduction du nombre d'itérations nécessaires à l'achèvement de l'ASBR et de la consommation de mémoire (voir le tableau 4.3), réalisant ainsi une optimisation temporelle et spatiale.

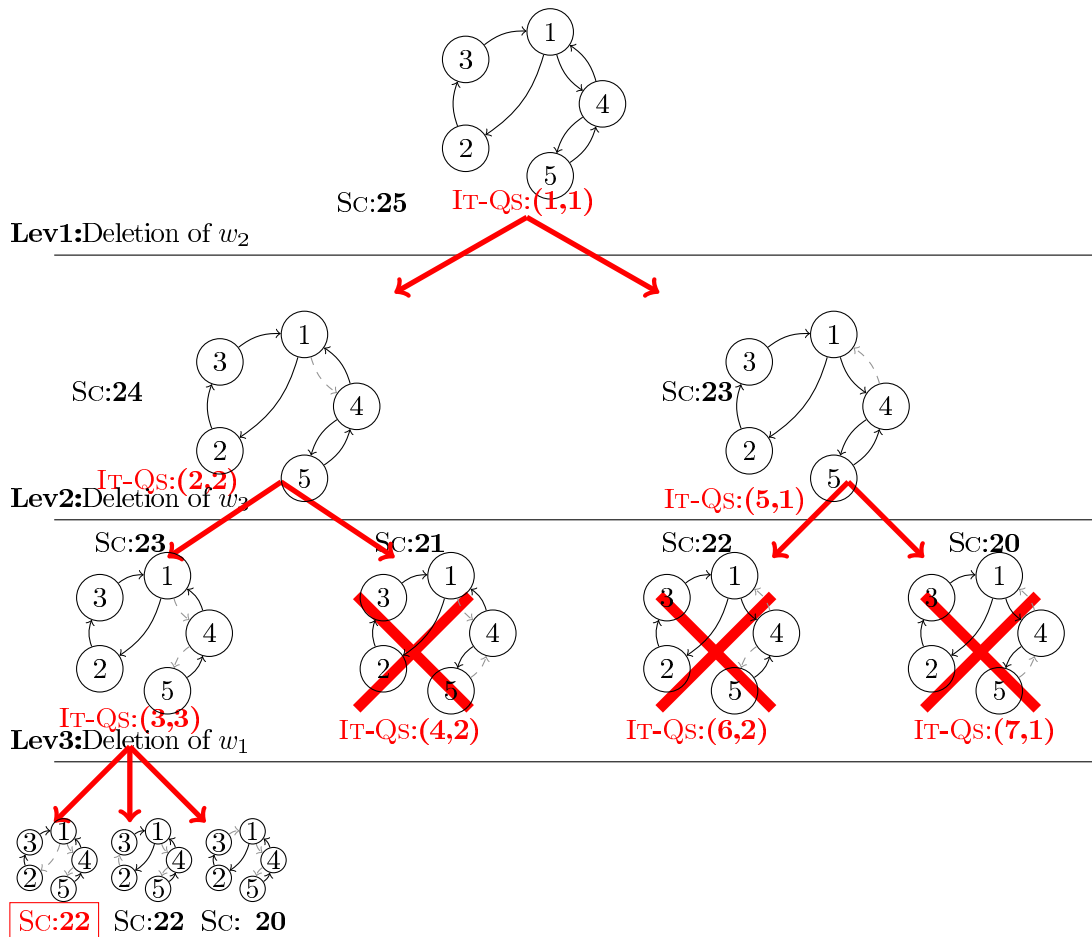


FIGURE 4.5 – Elimination des cycles selon les deux stratégies.

TABLE 4.3 – Comparaison entre la méthode Standard de B&B et la méthode B&B améliorée

	Taille max de la File	Nombre Itération	Itérations avant premier GOSC
Algorithme Standard B&B (Figure 4.2)	5	8	4
Première stratégie (Figure 4.3)	4	8	3
Deuxième stratégie (Figure 4.4)	4	7	4
Deux stratégies (Figure 4.5)	3	7	3

### 3.2.2 Algorithme Proposé

Le nouvel algorithme considère ces deux contributions, il se base sur la structure de données illustrée à la figure 4.6. Il comprend au maximum  $m$  files d'attente, où  $m$  représente le nombre maximal de cycles que peut contenir un graphe durant le processus d'apprentissage [93].

Chaque file contient  $N_i$  graphes, qui ont le même nombre de cycles  $i$ , et sont classés par ordre décroissant suivant ses scores. Les files d'attente sont classés par ordre croissant suivant le nombre de cycles.

Cette structure est similaire à une pile contenant des files en tant qu'éléments, où la priorité est donnée au graphe ayant le meilleur score, où ce graphe est sélectionné de la file qui contient l'ensemble des graphes ayant le moins nombre de cycles.

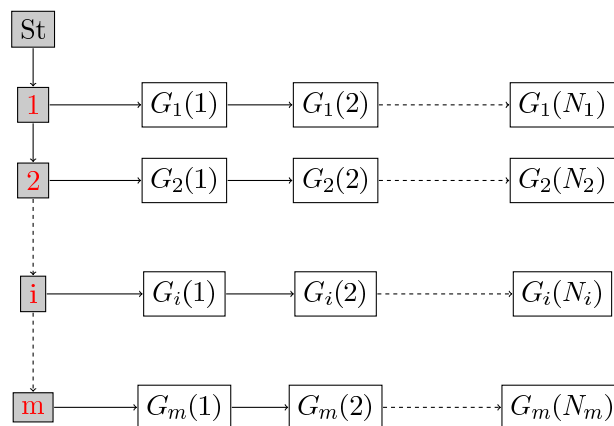


FIGURE 4.6 – Structure de donnée Utilisée

---

**Algo2** :Stratégie du parcours et divition

---

```

1 :  $S_{best} \leftarrow -\infty$ 
2 :  $Insert((G, H, S), St)$ 
3 : While NotEmpty(St) do
4 :      $(G_{cur}, H_{cur}, S_{cur}) \leftarrow take(St)$ 
5 :      $cyc \leftarrow ShortestCycle(G_{cur})$ 
6 :     For all arc  $(u \rightarrow v) \in cyc$  do
7 :          $Prohibited((u \rightarrow v), H_{cur})$ 
8 :          $(G, S) \leftarrow GetRestGraph(G_{cur}, H_{cur}, Cache(v, PA_v))$ 
9 :         If  $(S > S_{best})$  then
10 :              $NbCycr \leftarrow NumberCycle(G)$ 
11 :             If  $(DAG(G))$  then
12 :                  $S_{best} \leftarrow S$ 
13 :                  $G_{best} \leftarrow G$ 
14 :             Else
15 :                  $Insert((G, H_{cur}, S), St)$ 
16 :                  $MustExist((u \rightarrow v), H_{cur})$ 
17 :                  $Prohibited((v \rightarrow u), H_{cur})$ 
18 :             Endif
19 :         Endif
20 :     Endfor
21 : Enddo

```

---

Le nouvel algorithme **Algo2** qui améliore la version standard, se présente comme suit :

À la ligne 4, le triplet actuel  $(G_{cur}, H_{cur}, S_{cur})$  est choisit de la structure  $S_t$  pour l'exploration. On note que chaque triplet  $(G, H, S)$  est classé dans  $St$ , d'abord suivant le nombre de cycles de  $G$  par ordre croissant, puis suivant le score  $S$  par ordre décroissant (voir la figure 4.6) (la première contribution est appliquée).

Le cycle le plus court de  $G_{cur}$  est sélectionné à la ligne 5 (la deuxième contribution est appliqué). Le cycle choisi est supprimé, ce qui donne de nouveaux

graphes (ligne 6). Pour chaque graphe obtenu, le score  $S$  est calculé (ligne 8), puis vérifié s'il est supérieur au meilleur score obtenu jusqu'à présent  $S_{meilleur}$  (ligne 9). Si c'est le cas, la fonction  $NumberCycle(G)$  renvoie  $NbCycr$  (le nombre de cycles que contient  $G$ ) (voir ligne 10), alors on vérifie si  $G$  est un *GOSC*, c.-à-d. si  $NbCycr \leq 0$ , alors le meilleur score et le meilleur *GOSC* sont mis à jour aux lignes 12,13. Sinon, il est stocké dans la structure en fonction de son nombre de cycles  $NbCycr$  et de son score  $S$  (ligne 15) (la première contribution est appliquée).

À la ligne 9, si le score  $S$  du graphe est inférieure au meilleur score obtenu jusqu'à présent, le graphe est complètement retiré de l'arbre de recherche.

Cet algorithme s'arrête lorsque la structure  $St$  est vide (ligne 3).

Il faut noter aussi que dans l'algorithme standard, les *GOSCs* sont aussi enfilés dans  $Q$ , et sont récupérés par la suite de la file d'attente  $Q$  (voir ligne 4 algorithme standard), alors que dans le nouvel algorithme les *GOSCs* ne sont plus enregistrés dans la file, c.à.d. chaque graphe est vérifié si c'est un *GOSC* (voir ligne 11 nouvel algorithme), si c'est le cas le nouveau résultat est modifié si c'est nécessaire, sinon il sera enfilé. Donc la file d'attente utilisée par le nouvel algorithme ne contient pas des *GOSCs*. Cette procédure permet ainsi de minimiser la taille de la file.

### 3.2.3 Expérimentations

Après ces deux contributions ayant trait à l'optimisation de l'ASRB, nous avons besoin de valider ces approche par une série d'expérimentations. En plus des bases de données médicaux décrites dans l'annexe, nous avons ajouté d'autres bases de données dans divers domaines, contenant différents nombre de variables, disponibles sur le site *UCI* [39]. Les données manquantes sont supprimées et les variables continues sont discrétisées, voir tableau 4.4.

TABLE 4.4 – Bases de données utilisées dans les expérimentations

Bases de Données	Nombre de variables	Nombre d'instances
Car	7	1728
Nursery	9	12960
Flare	12	1066
Letter	17	20000
Bank	17	5000
Hepatitis	20	155
German Credit	21	1000
Cardiotocography	23	2126
Wall Following Robot Navigation (WFRN)	25	5456
KEGG Metabolic Reaction Network (KMRA)	28	65554
Wisconsin Diagnostic Breast Cancer (WDBC)	31	683

Le score *BIC* décrit à la section 1.8 est adopté comme critère de sélection du modèle.

TABLE 4.5 – Taille du Cache et de l'Espace de Recherche

Bases de Données	Taille du Cache	Taille de l'Espace de Recherche
Car	21	$2^9$
Nursery	42	$2^{14}$
Flare	110	$2^{30}$
Letter	217	$2^{62}$
Bank	139	$2^{38.1}$
Hepatitis	125	$2^{43.6}$
German Credit	60	$2^{29.6}$
Cardiotocography	203	$2^{62.58}$
WFRN	198	$2^{65.8}$
KMRA	129	$2^{46.7}$
WDBC	139	$2^{51}$

Le tableau 4.5 décrit les noms des bases de données utilisées, suivi par la taille du cache généré et de la taille de l'espace des structures possibles (voir section

1.8).

Nous avons implémenté la méthode standard *B&B* de nouveau, car le code disponible en ligne<sup>1</sup> ne fournit pas toutes les sorties nécessaires pour notre comparaison. Nous avons également implémenté notre version modifiée de l'algorithme *B&B*.

Le C est le langage de programmation utilisé dans notre implémentation. C'est un langage de bas niveau, il est l'un des langages les plus utilisés ; il permet de gérer l'allocation de la mémoire et donc de l'optimiser. De plus, grâce à l'utilisation de pointeurs, il est possible d'améliorer les performances de tout logiciel.

Il faut noter aussi, que l'implémentation de la méthode Branch&Bound nécessite la création d'un arbre. Tout comme les listes chaînées, les arbres sont composés d'éléments, appelés souvent des nœuds. Ils sont semblables aux listes chaînées par le fait que les éléments sont chaînés les uns avec les autres, à l'aide des pointeurs. Donc l'utilisation du langage C a permis une meilleure implémentation de notre algorithme.

L'environnement logiciel de notre système de travail est la plate-forme Microsoft sous le système d'exploitation Windows 7.

Le matériel utilisé est un Pentium 4 2.8 GHz avec 1Go de RAM.

L'algorithme B&B produit un grand nombre de solutions réalisables (Solutions Approximatives). Dans notre comparaison, nous nous basons que sur la première solution réalisable trouvée en plus de la solution optimale (voir Tableaux 4.6, 4.9). Nous notons également que *Fail* signifie que l'algorithme ne pourrait pas résoudre le problème dans un délai de deux heures.

---

1. <http://www.ecse.rpi.edu/~cvrl/structlearning.html>

TABLE 4.6 – Résultats (taille de la mémoire & nombre d’itérations) pour obtenir la première solution par le nouveau B&B (respectivement le standard B&B).

Bases de Données	Nouveau B&B		Standard B&B	
	Taille	Itérations	Taille	Itérations
Car	2	2	2	2
Nursery	6	5	107	13
Flare	826	549	687800695	27225
Letter	147	74	229680877	16130
Bank	25	19	<i>Fail</i>	<i>Fail</i>
Hepatitis	37	25	1429696215	36883
German Credit	8	9	31806123	6933
Cardiotocography	838	323	75736283	8461
WFRN	40	27	13929212318	114112
KMRA	373	1825	171465192	14655
WDBC	127	75	23196504	5162

TABLE 4.7 – Ratios comparatives pour obtenir la première solution

Bases de Données	ratio Nb iteration (std/nv)	ratio Taille mem (std/nv)
Car	1	1
Nursery	18	3
Flare	832688	50
Letter	1562455	218
Bank	//	//
Hepatitis	3864033	1475
German Credit	3975765	770
Cardiotocography	90377	26
WFRN	34822808	4226
KMRA	45969228	8
WDBC	182649	69

TABLE 4.8 – Ratios comparatives pour obtenir la solution optimale

Bases de Données	ratio Nb itereration (std/nv)	ratio Taille mem (std/nv)
Car	1	1
Nursery	17	3
Flare	//	//
Letter	1562455	218
Bank	//	//
Hepatitis	82506560	7
German Credit	3975765	770
Cardiotocography	1359096	3
WFRN	//	//
KMRA	261779	6
WDBC	182650	69

### 3.2.4 Discussions

TABLE 4.9 – Résultats (taille de la mémoire & nombre d’itérations) pour obtenir la solution optimale par le nouveau B&B (respectivement le standard B&B).

Datasets	Score	Nouveau B&B		Standard B&B	
		Taille	Itérations	Taille	Itérations
Car	-13102.6	9	5	11	5
Nursery	-126283	6	5	107	13
Flare	-6785.45	826	549	<i>Fail</i>	<i>Fail</i>
Letter	-603967	147	74	229680877	16130
Bank	-78887.5	169	27653	<i>Fail</i>	<i>Fail</i>
Hepatitis	-905.699	32	7593	2640209934	54115
German Credit	-16096.1	8	9	31806123	6933
Cardiotocography -	58908.8	125	4605	169886974	12690
WFRN	-324748	111	1287	<i>Fail</i>	<i>Fail</i>
KMRA	-6970.02	655	2550	171465192	14655
WDBC	-28567.5	127	75	23196504	5162

TABLE 4.10 – Temps d'exécution du nouveau *B&B* vs celui de différents algorithmes de l'état de l'art. Les temps sont donnés en secondes

Datasets	Nouveau B&B	Standard B&B	MEDP	A*	ratio Std / Nv
Car	0.001	0.001	0.296	0.281	1
Nursery	0.001	0.031	0.545	0.187	31
Flare	0.781	<i>Fail</i>	1.014	0.172	/
Letter	52.937	70.516	4.118	1.28	1
Bank	12.375	<i>Fail</i>	305.043	14.18	/
Hepatitis	4.781	2337.047	28.657	0.749	489
German Credit	0.265	22.141	57.611	2.402	84
Cardiotocography	6.610	9.250	26456.559	7.348	1
WFRN	3.813	<i>Fail</i>	2547.687	5.582	/
KMRA	52.438	505.485	<i>Fail</i>	65.301	10
WDDB	9.969	441.953	<i>Fail</i>	18.237	44

La figure 4.7 représente le nombre maximum de cycles que peut avoir un graphe, ainsi que le nombre maximum d'arcs que peut contenir un graphe. Il faut noter que dans le cas où les graphes contiennent des cycles de même taille, la deuxième procédure qui tente de supprimer le cycle le plus court devient insignifiante. Cette figure nous aide à justifier les résultats dans les cas où nos améliorations sont légèrement significatives. (les cycles ont la même taille ; le choix du cycle le plus court devient donc trivial) ; il montre également dans quels cas nos améliorations sont établies de manière significative (plusieurs cycles avec des tailles différentes).

Comme mentionné précédemment, l'avantage principal de l'algorithme *B&B* par rapport à d'autres, est sa capacité de produire des solutions approximatives avant la fin de l'exécution. En effet, si l'exécution est arrêtée, la méthode fournit la meilleure solution actuelle ; il est donc très important de montrer l'efficacité du nouveau *B&B* par rapport à celle de l'algorithme précédent [37], en termes de taille de la file (consommation de mémoire) et de nombre d'itérations, pour obtenir la première solution approximative ainsi que la meilleure solution.

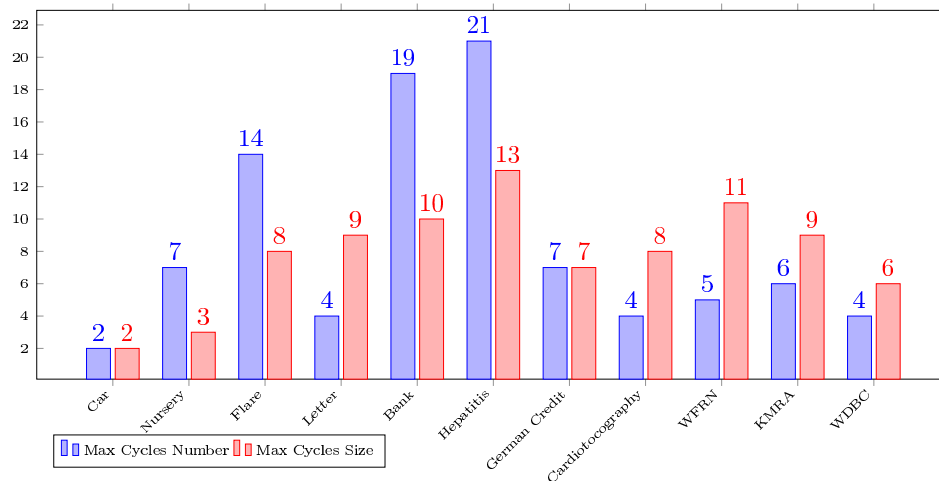


FIGURE 4.7 – Nombre et Taille maximal de cycles que peut contenir un graphe durant de l’exécution de l’algorithme.

Dans notre expérimentation, on a choisi deux petites bases de données (*Car*, *Nursery*) qui ont respectivement 7 et 9 variables. On remarque que la taille et le nombre de cycles sont limités par le nombre des variables contenu dans une base de donnée, c.-à-d. la taille du cycle ne dépasse jamais le nombre des variables.

Durant l’opération d’apprentissage de structure à partir de ces bases de données, le nouvel algorithme produit une légère amélioration, car les cycles ont presque les mêmes tailles (voir figure 4.7). Pour la base de donnée *Car* le nombre maximum des cycles que peut avoir un graphe est 2 et la taille de tous les cycles est égal à 2, donc choisir le cycle le plus court n’a pas sens.

Soient les bases de données *Hepatitis*, *Bank*, *Flare* et *WFRN*, qui produisent un grand nombre de cycles avec différentes tailles.

On remarque que l’algorithme standard de *B&B* appliqué à ces 3 dernières bases d’exemples n’arrive pas à fournir la solution optimale, dans un délai de deux heures, malgré leurs nombres limités de variables, tandis que le nombre de cycles ainsi que leurs taille sont remarquables. Le nouvel algorithme produit la solution optimal dans moins de deux minutes.

Nous constatons que pour la base de données *Bank*, l’algorithme standard n’a même trouvé la première solution approximative, et pour les bases de données *Flare* et *WFRN*, les premières solutions ont été trouvées difficilement (voir 4.8). *Hepatitis* est parmi les bases de données, où l’algorithme standard a fourni la

solution optimale. En comparaison avec d'autres bases de données, le taux d'amélioration du nouvel algorithme appliqué à cette base de données est très élevé, (voir le tableau 4.10). Ce gain de temps est dû au nombre de cycles de différentes tailles que contient un graphe durant la phase d'apprentissage. (voir la figure 4.7)

L'application de l'algorithme sur la plus grande base de données (*WDBC* avec 31 variables), minimise le nombre d'itérations ainsi que la taille de mémoire, mais cette optimisation n'est pas significative en comparaison avec d'autres bases de données de taille inférieure, telle que *Letter* et *German Credit*, car ces dernières produisent durant la phase d'apprentissage plus de cycles avec une taille supérieure en comparaison avec la base de données *WDBC*.

Pour résumer cette comparaison avec la méthode standard *B&B*, notre stratégie améliore à la fois les coûts de calcul et la consommation de mémoire (le nombre d'itérations et le nombre de graphes stockés dans la structure de données sont considérablement réduits; voir Tableaux 4.6, 4.9), en particulier pour les graphes ayant plusieurs cycles avec différentes tailles.

Le tableau 4.10 indique que le temps d'exécution des quatre algorithmes distincts : la méthode *B&B* modifiée, l'algorithme *B&B* standard [37], Programmation dynamique avec efficacité de la mémoire (MEDP) [22], et l'algorithme de recherche  $A^*$  [23]. Les résultats obtenus par ce nouveau algorithme sont presque similaires à ceux obtenus par l'algorithme  $A^*$ , mais le nouveau algorithme est plus efficace (en raison de sa capacité de produire des solutions approximatives avant la fin de l'exécution); ces deux méthodes sont très rapides par rapport aux autres. Le plus mauvais algorithme en termes de performances est la méthode standard *B&B*.

Nous notons également que les deux plus grandes bases de données en nombre de variables (*KMRA* et *WDBC*) ne produisent pas de solution avec MEDP.

### 3.3 Stratégie de bornage

#### 3.3.1 Introduction

Nous décrivons dans cette section la troisième contribution [104], où nous expliquons le nouvel opérateur de calcul de la borne supérieure, et nous montrons son efficacité à l'aide des expérimentations.

Dans un arbre, pour chaque nœud (sous-problème) une borne supérieure est calculée. Si la valeur de cette borne est inférieure au score de la meilleure solution trouvée jusqu'à présent, le sous-problème est complètement élagué. Par conséquent, les performances d'un algorithme B&B dépendent essentiellement de la pertinence de stratégie utilisée pour le calcul de la borne supérieure. Dans l'algorithme standard B&B, le score  $S$  du graphe  $G$  est utilisé comme borne supérieure. Pour accélérer l'exploration d'arbres de recherche, nous utilisons un opérateur efficace pour le calcul de cette borne.

#### 3.3.2 Opérateur de bornage

Le nouvel opérateur de bornage est une fonction qui calcule la borne supérieure, cette borne est le score maximal des sous-graphes dérivés d'un graphe cyclique  $G$ . Les sous-graphes sont générés en éliminant l'ensemble  $C$  de Cycles Disjoints par leurs Sommet (CSD)(voir figure 4.8 Cycles colorés en rouge); tous les cycles de cet ensemble n'ont pas de sommet en commun.

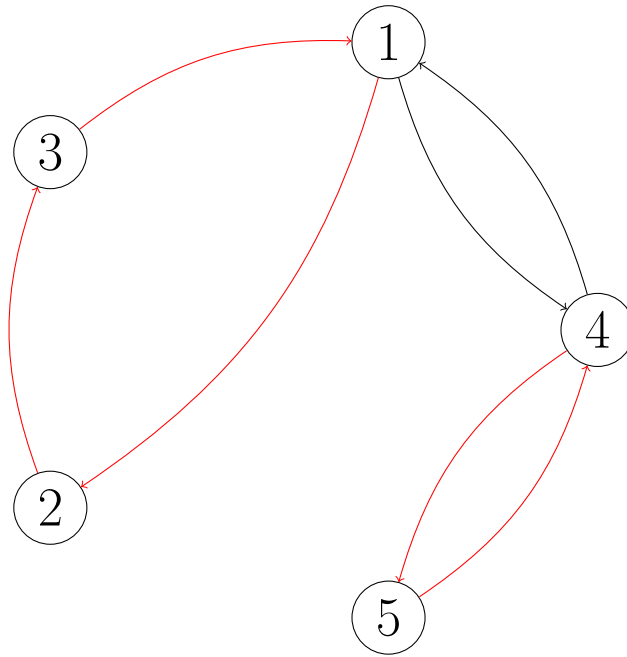


FIGURE 4.8 – Exemple d'un ensemble des Cycles Disjoint par leurs Sommet

Comme déjà décrit précédemment, l'algorithme *B&B* se base sur la suppression des cycles d'une manière répétitive jusqu'à l'obtention du meilleur GOSC. L'élimination d'un cycle contenant un nombre de nœuds (variables) passe par la suppression de tous ses arcs. A chaque suppression d'un arc ( $u \rightarrow v$ ), on aura comme résultat un sous-graphe. Le score de ce sous-graphe est recalculé en se basant sur le tableau du cache, où le score local de la variable  $v$  est modifié dans le score global, les parents de la variable  $v$  ne doit pas contenir  $u$  et doit répondre à des contraintes exigées par la matrice  $H$ .

L'algorithme de recherche de la borne supérieure est estimé en éliminant l'ensemble  $C$  de Cycles Disjoint par leurs Sommet (CSD). A chaque suppression d'un cycle, il calcule le meilleur score global, cette opération est répétée jusqu'à l'élimination de tous l'ensemble  $C$  de CSD. L'algorithme effectue au maximum  $N$  suppressions d'arc pour un graphe contenant  $N$  nœuds. Et pour chaque variable du cache la fonction l'utilise au plus une fois.

Alors que, dans le cas où nous avons des cycles avec des variables en commun (voir figure 4.9), la fonction peut utiliser la même variable plusieurs fois. Dans tel cas la fonction doit prendre plusieurs paramètre pour calculer cette borne, en plus la fonction doit aussi utiliser la matrice  $H$ , ce qui complique le traitement, et par conséquent son utilisation alourdi l'exécution de l'algorithme *B&B*.

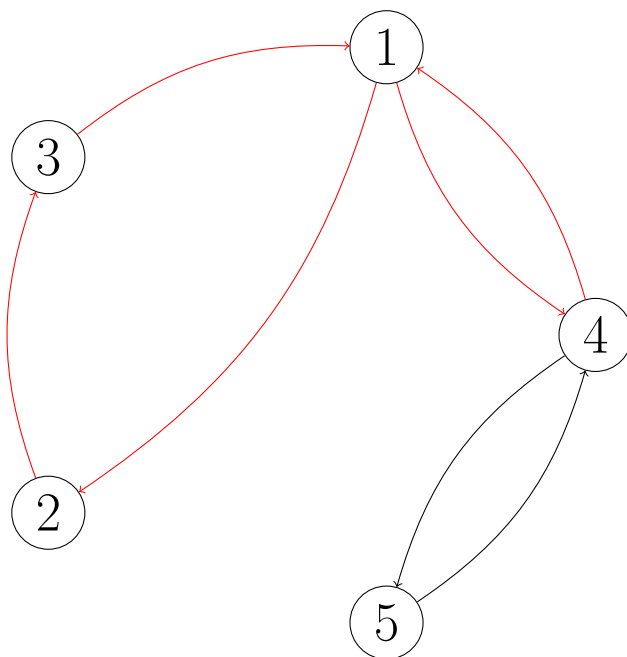


FIGURE 4.9 – Exemple deux Cycles avec un Sommet en commun

Il faut noter aussi que dans le cas où l'ensemble des cycles contient des arcs en commun, la suppression d'un arc parmi cet ensemble permet d'éliminer deux cycles ( voir figure 4.10 arc  $(1 \rightarrow 2)$ )

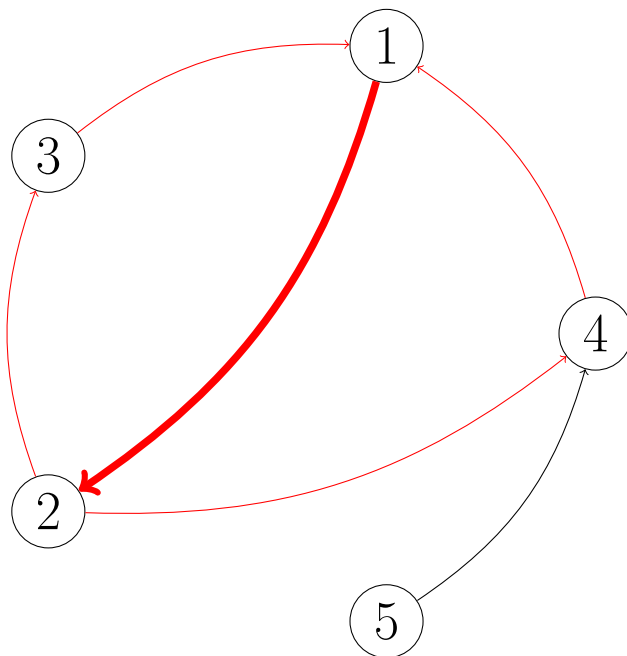


FIGURE 4.10 – Exemple deux Cycles avec un arc en commun

Nous considérons que cet ensemble de cycles, pour éviter l'explosion combinatoire. D'après la littérature, le processus de recherche de l'ensemble de CSD

répondant à certaines contraintes peut être un problème NP-difficile [99]. Dans notre contribution, aucune contrainte n'a été utilisée.

Avant d'entamer la fonction *UpBound*, nous allons définir la fonction *FindVDC*, qui permet de rechercher l'ensemble de *VDC*.

Soit un graphe *G* avec *N* variables, et *M* représente le nombre de tous les cycles dans le graphe *G*, ces cycles sont stockés dans une structure *CyclesG*, *Verif* est un vecteur de *N* valeurs booléennes, initialisées à *false*. Pour chaque nœud, un jalon de son état est indiqué par ce vecteur, qui peut être soit présent (*true*) ou absent (*false*). La fonction *FindVDC* est comme suit :

---

### Pseudocode of FindVDC (G)

---

```

1 :  $i \leftarrow 1$ 
2 :  $CycleCurrent \leftarrow CyclesG[i]$ 
3 : While(( $i \leq M$ ) && !all(Verif)) do
4 :   If(All node in  $CyclesG[i]$  are checked absent in  $Verif$ ) Then
5 :     For All nœud  $u$  in  $CyclesG[i]$  do
6 :        $Verif[u] \leftarrow True$ 
7 :     EndFor
8 :      $Add\ CyclesG[i]\ in\ VDC$ 
9 :   EndIf
10 :   $i \leftarrow i + 1$ 
11 : EndWhile
12 : Return  $VDC$ 
EndFunction

```

---

Soit  $(G_{up}, H_{up}, S_{up})$  le triplet courant, La fonction *UpBound* localise tout ensemble *C* de VDC dans le graphe  $G_{up}$  (applique la fonction *FindVDC*). En ligne 3, pour chaque cycle  $C_{cur}$  de l'ensemble *C*, il localise le meilleur sous-graphe obtenu lors de la suppression de ce cycle. Comme indiqué dans cette fonction, pour chaque arc  $(u \rightarrow v)$  dans  $C_{cur}$  :

---

**Pseudocode of UpBound(G,H,S)**

---

```

1 :  $(G_{up}, H_{up}, S_{up}) \leftarrow (G, H, S)$ 
2 : Find a set  $C$  of Vertex-Disjoint Cycles in the graph  $G_{up}$ 
3 : For all cycle  $C_{cur} \in C$  do
4 :    $BestScore \leftarrow -\infty$ 
5 :   For all arc  $(u \rightarrow v)$  in  $C_{cur}$  do
6 :     Prohibited( $(u \rightarrow v), H_{up}$ )
7 :      $(G_{up}, S_{up}) \leftarrow GetRestGraph(G_{up}, H_{up}, Cache(v, PA_v))$ 
8 :     If  $(S_{up} > BestScore)$  Then
9 :        $BestScore \leftarrow S_{up}$ 
10 :       $(x, y) \leftarrow (u, v)$ 
11 :     EndIf
12 :     MustExist( $(u \rightarrow v), H_{up}$ )
13 :   EndFor
14 :   Prohibited( $(x \rightarrow y), H_{up}$ )
15 :    $(G_{up}, S_{up}) \leftarrow GetRestGraph(G_{up}, H_{up}, Cache(y, PA_y))$ 
16 : EndFor
17 : Return  $S_{up}$ 
EndFunction

```

---

1. La ligne 6 indique sur  $H_{cur}$  que cet arc est interdit.
2. Le nouveau score du sous-graphe obtenu est calculé sur la base de la restriction du  $H_{cur}$ . Ceci est réalisé en utilisant le cache  $C(v, PA_v)$ , ce qui évite le recalcul de tous les scores locaux, ligne 7.
3. Le nouveau score est vérifié, s'il améliore le  $BestScore$  (ligne 8). Si c'est le cas,  $BestScore$  et  $BestArc$  sont mis à jour aux lignes 9 et 10.
4. À la ligne 12, l'arc  $(u \rightarrow v)$  est marqué dans  $H_{cur}$  dans le graphe qu'il doit exister, puis l'algorithme passe à l'arc suivant.
5. Le doublet  $(G_{up}, S_{up})$  est mis à jour avec le meilleur graphe obtenu lors de l'élimination du cycle en cours, lignes 14 et 15.

A la fin, la fonction retourne la borne supérieure  $S_{up}$ .

### 3.3.3 Algorithme proposé

---

#### Algorithm 3 : Pseudocode of CB-B&B Method

---

```

1 :  $S_{best} \leftarrow -\infty$ 
2 : Insert( $(G, H, S, Up), Q$ )
3 : While NotEmpty( $Q$ ) do
4 :  $(G_{cur}, H_{cur}, S_{cur}, Up_{cur}) \leftarrow \text{take}(Q)$ 
5 : If ( $Up_{cur} > S_{best}$ ) then
6 :   If ( $S_{cur} > S_{best}$ ) then
7 :     If (DAG( $G_{cur}$ )) then
8 :        $S_{best} \leftarrow S_{cur}$ 
9 :        $G_{best} \leftarrow G_{cur}$ 
10 :    Else
11 :       $cyc \leftarrow \text{ShortestCycle}(G_{cur})$ 
12 :      For all arc  $(u \rightarrow v) \in cyc$  do
13 :        Prohibited( $(u \rightarrow v), H_{cur}$ )
14 :         $(G, S) \leftarrow \text{GetRestGraph}(G_{cur}, H_{cur}, \text{Cache}(v, PA_v))$ 
15 :         $Up \leftarrow \text{UpBound}(G_{cur}, H_{cur}, S)$ 
16 :        Insert( $(G, H_{cur}, S, Up), Q$ )
17 :        MustExist( $(u \rightarrow v), H_{cur}$ )
18 :        Prohibited( $(v \rightarrow u), H_{cur}$ )
19 :      Endfor
20 :    Endif
21 :  Endif
22 :
23 : Enddo

```

---

Le nouvel algorithme nommé *CB – B&B* améliore la méthode standard de *B&B*, où la stratégie du branchement décrite précédemment (voir section 3) a été

adopté, ainsi que le nouvel opérateur du bornage, qui a pour objective de calculer la borne supérieure. Les notations dans cet algorithme sont comme suit :

- $U_p$  : est la borne supérieure d'un sous-problème.
- $Q$  : est une file d'attente prioritaire de quadruplets  $(G, H, S, U_p)$ , classés en fonction de la valeur décroissante de  $S$  ou de la borne supérieure  $U_p$ .

L'objectif principal de cette méthode est de maximiser une fonction de score décomposable. Nos améliorations apportées à l'algorithme sont comme suit :

1. Le quadruplet  $(G_{cur}, H_{cur}, S_{cur}, U_{p_{cur}})$  est pris de la file  $Q$  (ligne 4).
2. À la ligne 5, nous vérifions si la borne supérieure actuelle  $U_{p_{cur}}$  est meilleure que  $S_{meilleur}$ . Si tel est le cas, il passe à l'instruction suivante. Sinon, sa branche associée est complètement éliminée.
3. A la ligne 11, le cycle le plus court sera choisi pour être éliminé (*La stratégie du branchement est appliqué*).
4. On insère le quadruplet  $(G, H_{cur}, S, U_p)$  dans  $Q$ , en fonction de son score  $S$  ou en fonction de sa borne supérieure  $U_p$ , ligne 16.

### 3.3.4 Expérimentations

Dans cette section, nous présentons des expérimentations visant à évaluer les performances de l'algorithme  $CB - B\&B$ , où nous comparons ses résultats avec d'autres algorithmes décrits dans l'état de l'art.

Ces études expérimentales ont été effectuées sur quatre ensembles de données *mdicales* Nursery, Hepatitis, Cardiotocography, Wisconsin Diagnostic Breast Cancer (WDBC) disponibles dans le répertoire *UCI* [39]. Les données manquantes sont supprimées et les variables continues sont discrétisées. Dans toutes les expériences, il n'y avait aucune contrainte et le score  $BIC$  est adopté comme critère de sélection du modèle.

Toutes les expériences sont réalisées sur le cluster *IG Computing de Mons Polytechnique*. Le langage  $C$  est utilisée pour tous les programmes.

TABLE 4.11 – Comparaison entre  $CB - B\&B$  et  $B\&B$  en termes de mémoire

Datasets	B&B	CB-B&B (Bran.Str.)	CB-B&B (Bran.&Bnd.Str.) (Score ordered)	CB-B&B (Bran.&Bnd.Str.) (Up ordered)
Nurse.	21	20	8	4
Hepat.	3848	1003	778	581
Cardio.	4874	801	452	508
WDBC	4589	2941	2399	2150

TABLE 4.12 – Comparaison entre  $CB - B\&B$  et  $B\&B$  en termes de nombre d'itérations

Datasets	B&B	CB-B&B (Bran.Str.)	CB-B&B (Bran.&Bnd.Str.) (Score ordered)	CB-B&B (Bran.&Bnd.Str.) (Up ordered)
Nurse.	67	58	10	10
Hepat.	8172	2682	1396	900
Cardio.	22243	3528	2258	2193
WDBC	21657	14521	10534	10534

Dans ce qui suit (Tableaux 4.11,4.12), nous montrons les comparaisons entre :

- **CB-B&B (Bran.Str.)** : qui applique uniquement la stratégie de branchement,
- **CB-B&B (Bran.&Bnd.Str. S ordered)** : Application des stratégies de branchement et de bornage et les sous-problèmes sont classés dans la file suivant  $S$ .
- **CB-B &B (Bran.&Bnd.Str. Up ordered)** : Application des stratégies de branchement et de bornage et les sous-problèmes sont classés dans la file suivant  $Up$ .

### 3.3.5 Discussions

La performance de tout algorithme  $B\&B$  dépend principalement de la pertinence de l'opérateur de bornage utilisé. Dans ce travail, la borne supérieure est représentée par le score maximum de sous-graphes obtenus, lorsque nous éliminons tout ensemble de CSD. Cela permet de résoudre le problème en moins d'itérations (voir Tableaux 4.11,4.12).

Il est possible d'élaguer tous les sous-problèmes dont la valeur de la borne supérieure ne dépasse pas la valeur de la borne inférieure (La valeur du meilleur score trouvé jusqu'à présent), car on peut affirmer que la solution optimale globale ne peut être contenue dans le sous-ensemble de solution représenté par ce sous-problème.

En ce qui concerne le stockage des quadruplets dans la file, Le classement se fait suivant la borne supérieure au lieu du score ; ce procédé améliore légèrement la performance de la nouvelle méthode.

Le tableau 4.13 indique le temps d'exécution de quatre algorithmes distincts : notre algorithme *CB – B&B* [104] , l'algorithme *B&B* standard [37], Programmation dynamique avec efficacité de la mémoire (*MEDP*) [22], et l'algorithme de recherche *A\** [23].

TABLE 4.13 – Temps d'exécution de *CB-B&B* par rapport aux différents algorithmes de l'état de l'art. Le temps d'exécution est en millisecondes (*Fail* signifie que la méthode n'a pas pu résoudre le problème dans un délai de deux heures) .

Base	CB-B&B	B&B		MEDP		A* Method	
	Time	Time	Taux	Time	Taux	Time	Taux
Nurse.	10	39	3.90	55.28	5.53	286.64	28.66
Hepat.	11	72	6.55	5.25	0.48	7.42	0.67
Cardio.	285	2417	8.48	43223.07	151.66	470.89	1.65
WDBC	1792	3957	2.21	Fail	—	4625.64	2.58
<b>Moyenne</b>	—	—	3.09	—	27.51	—	2.57

Pour chacune de ces trois dernières méthodes, nous utilisons un taux qui compare les résultats obtenus par chaque méthode avec la méthode *CB – B&B*. En comparaison avec ces techniques, la méthode proposée est très efficace, elle est en moyenne 3,09 fois plus rapide que la version standard *B&B* (Tableau 4.13). Elle présente aussi un avantage important : elle pourrait être arrêtée à tout moment ; dans ce cas, elle est considérée comme une heuristique et fournit la meilleure solution trouvée, jusqu'au moment d'arrêt.

Notez que tous ces algorithmes sont meilleurs que la méthode *MEDP*, qui

n'arrive pas résoudre les problèmes complexes (voir *WDBC* avec 31 variables dans le tableau 4.13).

### 3.4 Parallel-CB-B&B Algorithm

#### 3.4.1 Introduction

Cette section décrit notre quatrième contribution [104]. Nous présentons la version parallèle de l'algorithme *CB – B&B*, dans lequel la procédure principale est nommée *Parallel – CB – B&B – Main* et la procédure *Parallel – CB – B&B – Thread* est un thread de CB-B&B.

#### 3.4.2 Algorithme parallèle

Tous les threads exécutent *Parallel – CB – B&B – Thread* et partagent les variables suivantes :

1.  $Q$  est la file de quadruplet  $(G, H, S, Up)$  qui contient initialement un simple triplet avec  $G, H, S$  et  $Up$  comme borne supérieure de  $G$ ,
2.  $S_{best}$  enregistre la meilleure solution trouvée.

Deux variables d'exclusion mutuelles sont utilisées (*mutex- $Q$*  et *mutex- $S_{best}$* ) pour fournir un accès exclusif entre les threads, pour les variables partagées  $Q$  et  $S_{best}$ . Pour la lisibilité de l'algorithme, ces variables n'apparaissent pas dans l'algorithme ci-dessus.

---

#### Procédure Parallel-CB-B&B-Main

```

1:  $S_{best} \leftarrow -\infty$ 
2: Insert( $(G, H, S, Up), Q$ )
3: For  $i = 1 \rightarrow NBR - THREADS$  do
4:   thread-run(Parallel-B&B-Thread)
5: Endfor
EndProcédure

```

---

---

**ProcedureParallel-CB-B&B-Thread**

```

1: While (True) do
2:  $cond1 \leftarrow$  (All threads are waiting)
3:  $cond2 \leftarrow$  ( $Q$  is empty)
4: If ( $cond1 \&\& cond2$ ) then
5:   Stop program
6: Endif
7: If Empty( $Q$ ) then
8:    $Current - Thread - State \leftarrow Wait$ 
9: Else
10:  ( $G_{cur}, H_{cur}, S_{cur}, Up_{cur}$ )  $\leftarrow$  take( $Q$ )
11:  If ( $Up_{cur} > S_{best}$ ) then
12:    If ( $S_{cur} > S_{best}$ ) then
13:      If (DAG( $G_{cur}$ )) then
14:         $S_{best} \leftarrow S_{cur}$ 
15:         $G_{best} \leftarrow G_{cur}$ 
16:      Else
17:         $cyc \leftarrow$  ShortestCycle( $G_{cur}$ )
18:        For all arc ( $u \rightarrow v$ )  $\in$   $cyc$  do
19:          Prohibited( $(u \rightarrow v), H_{cur}$ )
20:          ( $G, S$ )  $\leftarrow$  GetRestGraph( $G_{cur}, H_{cur}, Cache(v, PA_v)$ )
21:           $Up \leftarrow$  UpBound( $G_{cur}, H_{cur}, S$ )
22:          Insert( $(G, H_{cur}, S, Up), Q$ )
23:          MustExist( $(u \rightarrow v), H_{cur}$ )
24:          Prohibited( $(v \rightarrow u), H_{cur}$ )
25:        Endfor
26:      Endif
27:    Endif
28:  Endif
29: Enddo

```

---

Cet algorithme insère le graphe initial représenté par le quadruplet  $(G, H, S, Up)$  dans  $Q$  et la valeur initiale de  $S_{best}$  prend  $-\infty$ .

Dans la première procédure NBR-THREADS, les threads sont lancés. Chacun de ces threads est une exécution de la procédure de Parallel-CB-B&B-thread. Dans chaque itération de cette procédure, il est vérifié si la résolution du problème est terminée ; l'algorithme se termine *ssi* les conditions *cond1* et *cond2* sont vérifiées (voir Ligne 4). La première condition indique que tous les threads sont en attente, la seconde signale que la file d'attente *Q* est vide.

À la ligne 7, *Q* est vérifiée, si elle est vide ; si c'est le cas, l'état du thread actuel passe à *Wait*, il doit attendre jusqu'à ce qu'un sous-problème soit ajouté à *Q* ou que la première condition *cond1* soit vérifiée. Si *Q* devient non vide, le thread actuel bascule son état vers *Actif*, prend le sous-problème de *Q* et le résout.

Les étapes restantes de cet algorithme sont expliquées dans l'algorithme précédent.

### 3.4.3 Expérimentations

Pour évaluer la performance de la méthode *Parallel CB – B&B*, des études expérimentales ont été menées uniquement sur trois grands ensembles de données médicales, à savoir *Hepatis*, *Cardiotocography* et *WDBC*. Nous avons ignoré la base de donnée *Nursery*, cette base contient un ensemble limité de données, par conséquent le degré de parallélisme ne peut pas être important.

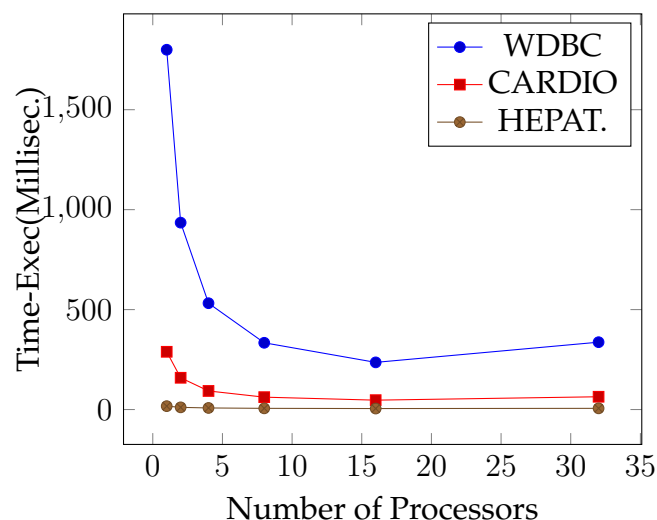


FIGURE 4.11 – Temps d'exécution de la méthode *Parallel CB–B&B*, pour l'ASRB appliqué aux bases de données médicales.

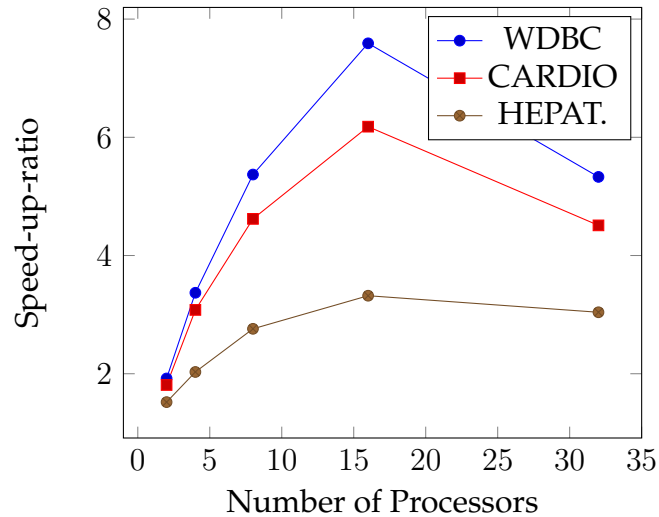


FIGURE 4.12 – Taux d’accélération de la méthode *Parallel CB – B&B*, pour l’ASRB appliqué aux bases de données médicales.

Les résultats de la simulation sont donnés pour 1, 2, 4, 8, 16 et 32 threads (respectivement), ainsi que les résultats analytiques. La moyenne des temps de calcul est prise pour chaque *Base de Donne* à partir de plusieurs expériences.

La figure 4.12 montre les taux d’accélération de la méthode *Parallel CB – B&B* pour les ensembles de données médicales. Nous observons que les taux d’accélération augmentent avec l’augmentation du nombre de processeurs.

### 3.4.4 Discussions

Le meilleur temps d’exécution est fourni avec 16 threads (Figure 4.11). Ce résultat montre que le parallélisme de *CB – B&B* pour l’ASRB est évolutif.

Les meilleurs résultats sont fournis avec 16 threads. Nous remarquons également que l’accélération obtenue pour le plus grand ensemble de données *WDBC* est supérieure à celle des autres ensembles de données .

Nous notons que dans presque tous les ensembles de données, les taux d’accélération pour 32 processeurs ont été réduits par rapport à 16 processeurs. Donc, davantage de processeurs ne réduisent pas nécessairement le temps d’exécution. En effet, avec un grand nombre de processeurs, nous avons plus de parallélisme et donc plus de communications qui ont un impact sur le temps d’exécution.

## 4 Conclusions

Selon la littérature, la recherche de la meilleure structure d'un RB est un problème NP-difficile, à cause du nombre exponentiel de structures possibles, étant donné une base d'exemple avec un ensemble limité de variable. Pour cela, nous améliorons une méthode de recherche exacte qui est l'algorithme Branch&Bound. Au début du chapitre nous avons montré les limitations de l'algorithme B&B standard pour l'ASBR, où nous avons ciblé les stratégies du parcours et du branchement, ainsi que l'opérateur du bornage. Nous avons utilisés un exemple illustratif de l'ASBR pour une meilleure explication des limitations de cet algorithme, cet exemple nous a permis aussi d'expliquer nos deux premières contributions (stratégies parcours et branchement), où la première contribution consiste à utiliser une combinaison de deux stratégies de parcours (la "*recherche en profondeur*" et "*la recherche du meilleur d'abord*" ) et la deuxième tente de choisir le graphe qui produit le moins nombre de sous-graphes. D'après les expérimentations sur multiple bases de données, ces contributions améliorent considérablement l'algorithme standard.

Après nous avons expliqué le nouvel opérateur de bornage, l'efficacité de la méthode B&B dépend généralement de la pertinence de cet opérateur, il permet de produire une borne supérieure meilleure que celle utilisée par la méthode standard. L'algorithme B&B se base sur cette borne pour garder dans l'arbre de recherche que les graphes qui peuvent contenir la solution optimale. Nous avons justifié notre choix de cet opérateur par des figures et des explications. L'algorithme proposé dans cette partie contient à la fois la stratégie de branchement et l'opérateur de bornage. Nous avons utilisés d'autres expérimentations sur des bases de données médicales, pour montrer l'efficacité de ce nouvel algorithme. Finalement, nous avons implémenté ce dernier algorithme sur une plate forme parallèle. Les simulations ont montrés que les meilleurs résultats sont fournis avec 16 threads. Ces résultats ont été justifiés par gain du temps d'exécution.

# Conclusion Générale

Dans cette thèse nous avons traité le problème d'apprentissage de structure des réseaux bayésiens. Nous avons amélioré l'algorithme *Branch&Bound*, où, nous avons proposé quatre contributions.

La décision médicale se base sur des données prise des sources différentes, qui sont généralement incertaines et incomplètes. La modélisation peut être très utile pour réduire l'incertitude dans le raisonnement clinique. Parmi les techniques d'IA utilisées dans telle situation, le Réseau Bayésien constitue une alternative très intéressante, il apporte une meilleure modélisation à l'aspect incertain. Nous avons défini les propriétés de ce modèle bayésien ainsi que les différentes étapes pour sa construction. Ensuite nous avons présenté quelques scores utilisés pour mesurer la qualité de ses structures. Nous avons détaillé le score BIC utilisé durant notre expérimentation, où nous avons justifié notre choix. Par la suite, nous avons expliqué les différents algorithmes utilisés pour l'apprentissage de structure des réseaux bayésiens. Nous avons commencé par la description des méthodes approchés qui produisent des solutions réalisables qui ne sont pas toujours optimales. Ces méthodes peuvent être réparties en deux classes : les heuristiques et les métaheuristiques. Nous avons décrit brièvement les heuristiques les plus populaires tels que l'algorithme PC et l'algorithme MWST. Ensuite les méthodes exactes les plus populaires dans ce domaine tels que la programmation dynamique, la programmation lineaire et l'algorithme A\* ont été expliqué avec beaucoup de détails.

Plusieurs travaux ont été proposés dans le domaine médical. Nous nous sommes fixés l'implication des réseaux bayésiens dans le processus décisionnel. Ces techniques consistent à la construction des modèles à partir des données requises et des experts, pour aider le médecin à prendre des meilleures décisions.

Dans le dernier chapitre, nous avons détaillé l’algorithme B&B standard pour l’ASRB à travers un exemple illustratif ; cet algorithme commence par un graphe créé à partir de la meilleure configuration de toutes les variables (généralement c’est un graphe cyclique), et à chaque itération il supprime un cycle, jusqu’à l’obtention du meilleur GOSC. Après avoir montré ses limitations, nous avons ciblé les stratégies du parcours et de branchement, ainsi que la borne supérieure. Par la suite, nos contributions ont été exposées. La première consiste à donner la grande priorité d’exploration au graphe ayant le meilleur score parmi ceux qui ont le moins nombre de cycles. La seconde contribution élimine le cycle le plus court. La troisième contribution consiste à utiliser un opérateur de bornage, qui produit une borne supérieure inférieure à celle utilisé par l’algorithme standard. La dernière contribution consiste à implémenter l’algorithme Branch&Bound sur une plate-forme parallèle.

Nous avons validé nos contributions avec des expérimentations sur un ensemble de bases de données contenant divers nombre de variables, où, nous avons comparé nos propositions avec l’algorithme standard et d’autres algorithmes exacts décrits dans la littérature.

## Perspectives

Dans la continuité directe de notre travail de thèse, nous pouvons proposer plusieurs pistes de recherche nécessitant plus d’attention.

- Dans notre travail nous avons utilisé le score BIC pour mesurer la qualité d’une structure, néanmoins il serait aussi intéressant d’évaluer les structures avec une combinaison de différentes mesures.
- L’application de la nouvelle méthode B&B pour l’apprentissage de structure avec des contraintes multiples, ces contraintes peuvent être issues des experts du domaine médical ou des ontologies médicaux spécifiques dans le domaine étudié.

- La problématique des données manquantes n’a pas été traitée dans ce travail. Nous envisageons la poursuite des travaux de recherche en appliquons la méthode B&B pour l’ASRB en réalisant un prétraitement sur les données manquantes.
- Une piste intéressante pourrait alors consister à combiner cette méthode avec d’autres algorithmes évolutionnaires, ce procédé peut soutenir la phase d’apprentissage.
- La combinaison des réseaux bayésiens avec d’autres méthodes intelligentes en profitant des avantages des différentes méthodes existantes dans la littérature, ça permet d’améliorer la décision médicale.
- En plus de l’apprentissage de structure, l’inférence bayésien est aussi un problème Np-Difficile. Nous envisageons plus de concentrations dans nos futures recherches sur l’inférence bayésienne pour construire un système de décision optimisé.
- L’application de la théorie des possibilités dans le domaine médicale est envisagé dans notre perspective. Cette nouvelle théorie permet aussi une meilleure modélisation de l’incertitude.

# Bibliographie

- [1] DELAPLACE, Alain. Approche évolutionnaire de l'apprentissage de structure pour les réseaux bayésiens. 2007. PhD Thesis. Tours.
- [2] Pearl, J. (1997). Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference. Morgan Kaufmann, 2nd edition.
- [3] Verma, T. et Pearl, J. (1990). Equivalence and synthesis of causal models. In Proceedings of the Sixth Conference on Uncertainty and Artificial Intelligence, pages 220–227.
- [4] Neapolitan, R. E. (2004). Learning Bayesian Networks. Prentice Hall.
- [5] Pearl, J. (2000). Causality : Models, Reasoning, and Inference. Cambridge, England : Cambridge University Press.
- [6] Geiger D, Verma T, Pearl J. d-separation : From theorems to algorithms. In Machine Intelligence and Pattern Recognition 1990 Jan 1 (Vol. 10, pp. 139-148). North-Holland.
- [7] Chan, C-C., Celai Batur, and Arvind Srinivasan. " Determination of quantization intervals in rule based model for dynamic systems." Systems, Man, and Cybernetics, 1991.'Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference on. IEEE, 1991.
- [8] Holte, Robert C. "Very simple classification rules perform well on most commonly used datasets." Machine learning 11.1 (1993) : 63-90.
- [9] COLOT, O., et al. Information criteria and abrupt changes in probability laws. Signal Processing VII : Theories and Applications, 1994, 1855-1858.
- [10] Bozdogan, Hamparsum. "Akaike's information criterion and recent developments in information complexity." Journal of Mathematical Psychology 44.1 (2000) : 62-91.

- [11] Bouhamed, Heni, et al. A new learning structure heuristic of Bayesian networks from data. *Machine Learning and Data Mining in Pattern Recognition* (2012) : 183-197.
- [12] Chickering, D. M., Heckerman, D., & Meek, C. (2004). Large-sample learning of Bayesian networks is NP-hard. *The Journal of Machine Learning Research*, 5, 1287-1330.
- [13] Jie Cheng, David Bell, and Weiru Liu. An algorithm for bayesian network construction from data. In *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics AI&STAT'97*, pages 83–90, 1997.
- [14] Jie Cheng, David Bell, and Weiru Liu. Learning belief networks from data : An information theory based approach. In *Proceedings of the sixth ACM International Conference on Information and Knowledge Management CIKM*, pages 325–331, 1997.
- [15] Jie Cheng, Russell Greiner, Jonathan Kelly, David Bell, and Weiru Liu. Learning Bayesian networks from data : An information-theory based approach. *Artificial Intelligence*, 137(1–2) :43–90, 2002.
- [16] Judea Pearl. *Causality : Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, England, 2000.
- [17] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, prediction, and search*. Springer-Verlag, 1993.
- [18] Judea Pearl and Tom Verma. A theory of inferred causation. In James Allen, Richard Fikes, and Erik Sandewall, editors, *KR'91 : Principles of Knowledge Representation and Reasoning*, pages 441–452, San Mateo, California, 1991. Morgan Kaufmann.
- [19] CHOW, C. ; LIU, Cong. Approximating discrete probability distributions with dependence trees. *IEEE transactions on Information Theory*, 1968, 14.3 : 462-467.
- [20] Chickering, D. (1995). A transformational characterization of bayesian network structures. In Hanks, S. et Besnard, P., editors, *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pages 87–98.
- [21] Cooper, Gregory F., and Edward Herskovits. "A Bayesian method for the induction of probabilistic networks from data." *Machine learning* 9, no. 4 (1992) : 309-347.

- [22] Singh, Ajit P., and Andrew W. Moore. Finding optimal Bayesian networks by dynamic programming. Carnegie Mellon University. Center for Automated Learning and Discovery, 2005.
- [23] Yuan, Changhe, Brandon Malone, and Xiaojian Wu. "Learning optimal Bayesian networks using A\* search." In IJCAI proceedings-international joint conference on artificial intelligence, vol. 22, no. 3, p. 2186. 2011.
- [24] Jaakkola, Tommi, David Sontag, Amir Globerson, and Marina Meila. "Learning Bayesian network structure using LP relaxations." In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 358-365. 2010.
- [25] Grötschel, Martin, Michael Jünger, and Gerhard Reinelt. "On the acyclic subgraph polytope." *Mathematical Programming* 33, no. 1 (1985) : 28-42.
- [26] Cussens, James, Matti Järvisalo, Janne H. Korhonen, and Mark Bartlett. "Bayesian network structure learning with integer programming : Polytopes, facets and complexity." *Journal of Artificial Intelligence Research* 58 (2017) : 185-229.
- [27] COOPER, Gregory F. ; HERSKOVITS, Edward. A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, 1992, 9.4 : 309-347.
- [28] HECKERMAN, David ; GEIGER, Dan ; CHICKERING, David M. Learning Bayesian networks : The combination of knowledge and statistical data. *Machine learning*, 1995, 20.3 : 197-243.
- [29] BUNTINE, Wray. Theory refinement on Bayesian networks. In : Proceedings of the Seventh conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., 1991. p. 52-60.
- [30] HECKERMAN, David ; GEIGER, Dan ; CHICKERING, David M. Learning Bayesian networks : The combination of knowledge and statistical data. *Machine learning*, 1995, 20.3 : 197-243.
- [31] Schwarz, G. "Estimating the dimension of a model *The Annals of Statistics* 6 (2), 461-464." URL : <http://dx.doi.org/10.1214/aos/1176344136> (1978).
- [32] Akaike, Hirotugu. "A new look at the statistical model identification." *IEEE transactions on automatic control* 19.6 (1974) : 716-723.

- [33] R. R. Bouckaert, Bayesian belief networks : From construction to inference, doctoral dissertation, University Utrecht (2001).
- [34] Liu, Zhifa, Brandon Malone, and Changhe Yuan. "Empirical evaluation of scoring functions for Bayesian network model selection." *BMC bioinformatics* 13.15 (2012) : S14.
- [35] S. Oliver and S. Gholami, Locally consistent Bayesian network scores for multi-relational data, in *Proc. 26th Int. Joint Conf. Artificial Intelligence* (AAAI Press, 2017), pp. 2693–2700.
- [36] C. Lee and P. Van Beek, An experimental analysis of anytime algorithms for Bayesian network structure learning, in *Advanced Methodologies for Bayesian Networks* (PMLR, 2017), pp. 69–80.
- [37] De Campos, C. P., & Ji, Q. (2011). Efficient structure learning of Bayesian networks using constraints. *The Journal of Machine Learning Research*, 12, 663-689.
- [38] Teyssier, Marc, and Daphne Koller. "Ordering-based search : A simple and effective algorithm for learning Bayesian networks." *arXiv preprint arXiv :1207.1429* (2012).
- [39] Asuncion, A., & Newman, D. (2007). UCI machine learning repository.
- [40] *Approches de l'incertitude et son impact sur la décision*. Numéro 2011-06 des Cahiers de la Sécurité Industrielle, Institut pour une Culture de Sécurité Industrielle, Toulouse, France (ISSN 2100-3874).
- [41] Kahneman , Choices, values and frames, and A Tversky, Cambridge university press, 2000
- [42] ABECASSIS, Philippe. "Incertitude médicale et processus de décision." *Communication au Ateliers Thématiques en Économie de la Santé : Rationalité et processus de décision*, FORUM-LATEC, Dijon, Mars. 1999.
- [43] Higgs J, Jones M. Clinical reasoning in the health professions. In : Higgs J, Jones, M (Eds). *Clinical Reasoning in the Health Professions*. 2nd ed. Oxford : Butterworth-Heineman Ltd ; 2000 :3-14.
- [44] Harris I. New Expectations for Professional Competence. In : Curry L, Wegin, JF (Eds). *Educating Professionals Responding to new expectations for competence and accountability*. San Francisco : Jossey-Bath Publishers ; 1993 :17-52.

- [45] Newble D, Norman, van der Vleuten C. Assessing Clinical Reasoning. In : Higgs J, Jones M (Ed) Clinical Reasoning in the Health Professions. Oxford : Butterworth-Heinemann Ltd, 2000 :156-65
- [46] ZHOU, Xian, et al. Bayesian adaptive design for targeted therapy development in lung cancer—a step toward personalized medicine. *Clinical Trials*, 2008, 5.3 : 181-193.
- [47] CAUCHEMEZ, Simon, et al. A Bayesian MCMC approach to study transmission of influenza : application to household longitudinal data. *Statistics in medicine*, 2004, 23.22 : 3469-3487.
- [48] LANGMEAD, Christopher J. Generalized queries and bayesian statistical model checking in dynamic bayesian networks : Application to personalized medicine. 2009.
- [49] ZHU, Wen-Feng ; YAN, Jun-Feng ; HUANG, Bi-Qun. Application of Bayesian network in syndrome differentiation system of traditional Chinese medicine. *Zhong xi yi jie he xue bao= Journal of Chinese integrative medicine*, 2006, 4.6 : 567-571.
- [50] ARSENE, Octavian ; DUMITRACHE, Ioan ; MIHU, Ioana. Medicine expert system dynamic Bayesian Network and ontology based. *Expert Systems with Applications*, 2011, 38.12 : 15253-15261.
- [51] WIGGINS, M., et al. Evolving a Bayesian classifier for ECG-based age classification in medical applications. *Applied soft computing*, 2008, 8.1 : 599-608.
- [52] JIRIK, Radovan ; TAXT, Torfinn. Two-dimensional blind Bayesian deconvolution of medical ultrasound images. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 2008, 55.10.
- [53] FLORES, M. Julia, et al. Incorporating expert knowledge when learning Bayesian network structure : a medical case study. *Artificial intelligence in medicine*, 2011, 53.3 : 181-204.
- [54] ONIŚKO, Agnieszka ; DRUZDZEL, Marek J. Impact of precision of Bayesian network parameters on accuracy of medical diagnostic systems. *Artificial intelligence in medicine*, 2013, 57.3 : 197-206.

- [55] MCCORMICK, Tyler H., et al. Bayesian hierarchical rule modeling for predicting medical conditions. *The Annals of Applied Statistics*, 2012, 6.2 : 652-668.
- [56] ABRAHAM, Ranjit ; SIMHA, Jay B. ; IYENGAR, S. S. Medical datamining with a new algorithm for Feature Selection and Naïve Bayesian classifier. In : *Information Technology,(ICIT 2007). 10th International Conference on. IEEE, 2007. p. 44-49.*
- [57] RACZ, Michael J. ; SEDRANSK, J. Bayesian and frequentist methods for provider profiling using risk-adjusted assessments of medical outcomes. *Journal of the American Statistical Association*, 2010, 105.489 : 48-58.
- [58] FENTON, Norman ; NEIL, Martin. Comparing risks of alternative medical diagnosis using Bayesian arguments. *Journal of Biomedical Informatics*, 2010, 43.4 : 485-495.
- [59] CHEN, Yao-Tien. A level set method based on the Bayesian risk for medical image segmentation. *Pattern Recognition*, 2010, 43.11 : 3699-3711.
- [60] WAGHOLIKAR, Kavishwar B. ; VIJAYRAGHAVAN, Sundararajan ; DESHPANDE, Ashok W. Fuzzy naive Bayesian model for medical diagnostic decision support. In : *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE. IEEE, 2009. p. 3409-3412.*
- [61] CONSTANTINO, Anthony Costa, et al. From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support. *Artificial intelligence in medicine*, 2016, 67 : 75-93.
- [62] ABRAHAM, Ranjit ; SIMHA, Jay B. ; IYENGAR, S. Sitharama. Effective Discretization and Hybrid feature selection using Naïve Bayesian classifier for Medical datamining. *International Journal of Computational Intelligence Research*, 2009, 5.2 : 116-129.
- [63] NEELON, Brian ; O'MALLEY, A. James ; NORMAND, Sharon-Lise T. A Bayesian two-part latent class model for longitudinal medical expenditure data : assessing the impact of mental health and substance abuse parity. *Biometrics*, 2011, 67.1 : 280-289.

- [64] GADEWADIKAR, Jyotirmay, et al. Exploring Bayesian networks for medical decision support in breast cancer detection. *African Journal of Mathematics and Computer Science Research*, 2010, 3.10 : 225-231.
- [65] MANJUSHA, K. K. ; SANKARANARAYANAN, K. ; SEENA, P. Prediction of different dermatological conditions using naïve Bayesian classification. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2014, 4.1.
- [66] GUTMAN, Roe ; AFENDULIS, Christopher C. ; ZASLAVSKY, Alan M. A Bayesian procedure for file linking to analyze end-of-life medical costs. *Journal of the American Statistical Association*, 2013, 108.501 : 34-47.
- [67] BELCIUG, Smaranda ; GORUNESCU, Florin. Error-correction learning for artificial neural networks using the Bayesian paradigm. Application to automated medical diagnosis. *Journal of biomedical informatics*, 2014, 52 : 329-337.
- [68] BUCCI, Giacomo ; SANDRUCCHI, Valeriano ; VICARIO, Enrico. Ontologies and Bayesian networks in medical diagnosis. In : *System Sciences (HICSS)*, 2011 44th Hawaii International Conference on. IEEE, 2011. p. 1-8.
- [69] JUN-ZHONG, J. I., et al. A Bayesian network learning algorithm based on independence test and ant colony optimization. *Acta Automatica Sinica*, 2009, 35.3 : 281-288.
- [70] Zhang, Q., Li, Z., Zhou, C. J., & Wei, X. P. (2013). Bayesian network structure learning based on the chaotic particle swarm optimization algorithm. *Genetics and Molecular Research*, 12(4), 4468-4479.
- [71] Faulkner, E. (2007). K2GA : heuristically guided evolution of Bayesian network structures from data. In *2007 IEEE Symposium on Computational Intelligence and Data Mining*.
- [72] Dos Santos, E. B., Hruschka Jr, E. R., Hruschka, E. R., & Ebecken, N. F. (2010, July). A distance-based mutation operator for learning bayesian network structures using evolutionary algorithms. In *Evolutionary Computation (CEC), 2010 IEEE Congress on* (pp. 1-8). IEEE. .
- [73] Daly, R., & Shen, Q. (2009). Learning Bayesian network equivalence classes with ant colony optimization. *Journal of Artificial Intelligence Research*, 35(1), 391.

- [74] LARRAÑAGA, Pedro, et al. Structure learning of Bayesian networks by genetic algorithms : A performance analysis of control parameters. *IEEE transactions on pattern analysis and machine intelligence*, 1996, 18.9 : 912-926.
- [75] Vahid Rezaei Tabar. et al. Structure learning of Bayesian networks by genetic algorithms : A performance analysis of control parameters. *JIRSS* (2016) Vol. 15, No. 2, pp 45-61
- [76] STREET, W. Nick ; WOLBERG, William H. ; MANGASARIAN, Olvi L. In : *Biomedical Image Processing and Biomedical Visualization*. International Society for Optics and Photonics, 1993. p. 861-871.
- [77] MANDELBROT, Benoit B. *The fractal geometry of nature*. New York : WH freeman, 1982.
- [78] MISHRA, Krishna Mohan. *Application of Machine Learning Techniques to classify Fetal Hypoxia*. 2016.
- [79] GRIVELL, Rosalie M., et al. Antenatal cardiotocography for fetal assessment. *Cochrane Database of Systematic Reviews*, 2012, 12.
- [80] N. Chamidah and I. Wasito, "Fetal state classification from cardiotocography based on feature extraction using hybrid K-Means and support vector machine," in *International Conference on Advanced Computer Science and Information Systems, ICACISIS 2015*, October 10, Depok, Indonesia , pp. 37-41.
- [81] M. G. Signorini and G. Magenes, "Advanced signal processing techniques for CTG analysis," in *14th Mediterranean Conference on Medical and Biological Engineering and Computing, MEDICON 2016*, March 31 , Paphos, Cyprus, pp. 1199-1204.
- [82] V. Chudáček et al., "Open access intrapartum CTG database," *BMC Pregnancy Childbirth*, vol. 14, p. 16, 2014.
- [83] ZHANG, Yang ; ZHAO, Zhidong. Fetal state assessment based on cardiotocography parameters using PCA and AdaBoost. In : *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2017 10th International Congress on. IEEE, 2017. p. 1-6.

- [84] WINJE, B. A., et al. Analysis of 'count-to-ten' fetal movement charts : a prospective cohort study. *BJOG : an International Journal of Obstetrics & Gynaecology*, 2011, 118.10 : 1229-1238.
- [85] Heazell AE, Frøen JF. Methods of fetal movement counting and the detection of fetal compromise. *J Obstet Gynaecol* 2008 ;28 :147-54.
- [86] MACONES, George A., et al. The 2008 National Institute of Child Health and Human Development workshop report on electronic fetal monitoring : update on definitions, interpretation, and research guidelines. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, 2008, 37.5 : 510-515.
- [87] <https://geekymedics.com/how-to-read-a-ctg/#4>
- [88] STREET, P., et al. Short-term variation in abnormal antenatal fetal heart rate records. *American Journal of Obstetrics & Gynecology*, 1991, 165.3 : 515-523.
- [89] <http://www.operationalmedicine.org/ed2/Enhanced/LandD/Electronic-FetalMonitoring.htm>
- [90] KAMATH, R. S. ; KAMAT, R. K. Modelling Fetal Morphologic Patterns through Cardiotocography Data : A Random Forest based Approach. *RESEARCH JOURNAL OF PHARMACEUTICAL BIOLOGICAL AND CHEMICAL SCIENCES*, 2016, 7.5 : 2449-2455.
- [91] Y.BENMOUNA, M.A.CHIKH. Bayesian classifier for Breast cancer diagnosis. *BIOMEDICAL ENGINEERING INTERNATIONAL CONFERENCE (BIOMEIC'12) — OCTOBER 10-11,2012, TLEMEN (ALGERIA)*
- [92] Y.BENMOUNA, M.A.CHIKH. Application du classifieur bayésien pour la reconnaissance des arythmies cardiaques. 2ième Edition de la Conférence Nationale de l'Informatique Destinée aux Etudiants de Graduation et de Post-graduation(JEESI12) - 16 Avril, ALGER (ALGERIE)
- [93] Y.BENMOUNA, et al. New Method for Bayesian Network Learning. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 33, No. 02, 1959005 (2019)
- [94] DE CAMPOS, Luis M., et al. Ant colony optimization for learning Bayesian networks. *International Journal of Approximate Reasoning*, 2002, 31.3 : 291-311.
- [95] MAKOWSKI, David, et al. Prise en compte de l'incertitude en évaluation des risques : revue de la littérature et recommandations pour l'Anses. 2015.

- PhD Thesis. Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail.
- [96] PINCE, Marie. Gestion de l'incertitude en médecine générale : étude quantitative en Corrèze. 2013. PhD Thesis.
- [97] OKUDA, Tetsuji ; TANAKA, Hideo ; ASAI, K. A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. *Information and Control*, 1978, 38.2 : 135-147.
- [98] <https://www.futura-sciences.com/sante/dossiers/medecine-cancer-sein-1154/page/2/>
- [99] Krivelevich, Michael, et al. "Approximation algorithms and hardness results for cycle packing problems." *ACM Transactions on Algorithms (TALG)* 3.4 (2007) : 48.
- [100] Cung, V. D., et al. "Parallel and distributed branch-and-bound/A\* algorithms." Technical report, Laboratoire PRISM, Université de Versailles (1994) : 94.
- [101] Gendron, Bernard, and Teodor Gabriel Crainic. "Parallel branch-and-bound algorithms : Survey and synthesis." *Operations research* 42.6 (1994) : 1042-1066.
- [102] N. Melab. Contributions à la résolution de problèmes d'optimisation combinatoire sur grilles de calcul. LIFL, USTL, Novembre 2005. Thèse HDR.
- [103] Miller, Donald L., and Joseph F. Pekny. "Commentary—The Role of Performance Metrics for Parallel Mathematical Programming Algorithms." *ORSA Journal on Computing* 5.1 (1993) : 26-28.
- [104] Y.BENMOUNA, et al. Parallel Cycle-Based Branch and Bound Method for Bayesian network learning. *Pattern Analysis and Application*.(In Press) 2019
- [105] Naïm, P.,Wuillemin, P.-H., Leray, P., Pourret, O., et Becker, A. (2004). Réseaux bayésiens. Eyrolles, Paris.
- [106] CHAO, Yi-Sheng, et al. A network perspective of engaging patients in specialist and chronic illness care : The 2014 International Health Policy Survey. *PloS one*, 2018, 13.8 : e0201355.

# Annexe A

## Bases de données

### 1 Introduction

Pour valider nos contribution, nous avons utilisé 11 bases de données d'exemples disponibles sur le cite UCI [39]. Et puisque notre travail de recherche s'inscrit dans un cadre médicale, on a opté pour la description 3 bases de données médicales, ces dernières sont destinées au diagnostic médical, qui sont respectivement *Wisconsin Diagnostic Breast Cancer (WDBC)* utilisée pour la détection des cellules cancéreuses, *Cardiotography(CTG)* qui contient les informations collectés à partir d'un cardiogramme, c'est le moyen le plus utilisé pour évaluer le bien-être d'un fœtus, et la dernière base de donnée *Hepatitis* contient les caractéristiques de la maladie d'Hépatite, cette maladie peut provoquer un cancer du foie. Ces bases de données contiennent beaucoup plus d'attributs par rapport aux bases de données utilisées dans notre série d'expérimentations. Sachant que la base *WDBC* contient le plus grand nombre de variables(31),et d'après la littérature (voir section 4.5.2), les meilleures méthodes exactes connues pour l'ASRB sont appliquées sur un nombre de variables limitée (environ 30 variables). Cette base de donnée contient deux classes(Benin et maligne), Concernant la deuxième base de donnée CTG, elle contient 3 classes ; Normal, Suspect et Pathologique. Dans ce qui suit nous allons décrire en détails ces bases de données :

## 2 Le cancer du sein

### 2.1 Définition

Le cancer du sein est le cancer le plus répandu chez les femmes. D'après les statistiques [98], ce type de cancer représente le quart des nouveaux cas chez les femmes. Il provient de la dérégulation de certaines cellules qui se multiplient et forment le plus souvent une masse appelée tumeur. On distingue plusieurs types qui évoluent de différentes manières. Certains sont "dangereux" et évoluent très vite, d'autres plus lentement. Les cellules cancéreuses peuvent rester dans le sein, comme elles peuvent se propager à d'autres organes dans le corps humain, ce qui complique la situation. On parle alors de métastases. Généralement, la prolifération des cellules cancéreuses prend plusieurs mois, voire plusieurs années.

Après la détection d'une anomalie lors d'un examen de dépistage ou bien lors de la présence des symptômes chez une personne, plusieurs examens doivent être réalisés pour établir le diagnostic de cancer du sein, telle que l'examen anatomopathologique des tissus prélevés au niveau de la tumeur. Ce prélèvement est le plus souvent réalisé par micro ou macrobiopsies à travers la peau.

Plusieurs types de traitements sont indiqués pour traiter le cancer du sein : chirurgie, radiothérapie, hormonothérapie, chimiothérapie et thérapies ciblées. Dans certains cas, un seul type de traitement est nécessaire. Dans d'autres cas, une combinaison de traitements est essentielle pour mieux cerner la maladie. Par exemple, on peut effectuer une intervention chirurgicale puis terminer le traitement uniquement avec une chimiothérapie. Ces traitements peuvent causer le plus souvent des effets secondaires.

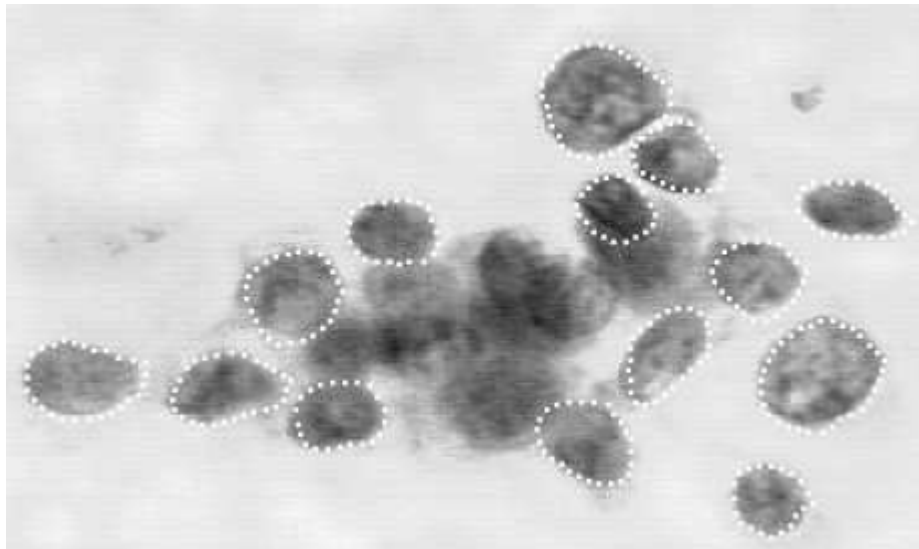
Le diagnostic des tumeurs du sein est traditionnellement réalisé par une biopsie complète ; une intervention chirurgicale invasive. L'aspiration à l'aiguille fine permet d'examiner une petite quantité de tissu de la tumeur. Cependant, le diagnostic avec cette procédure a rencontré un succès mitigé. En examinant attentivement les caractéristiques des cellules individuelles et des caractéristiques contextuelles importantes telles que la taille des masses cellulaires, les médecins de certaines institutions spécialisées ont pu diagnostiquer avec succès. Cependant, d'autres médecins ont constatés que de nombreuses caractéristiques différentes sont en corrélation avec la malignité, et le processus reste hautement subjectif, en

fonction des compétences et de l'expérience du médecin. Afin d'accroître la rapidité, la précision et l'objectivité du processus de diagnostic, des techniques de traitement d'image et d'apprentissage automatique sont fortement conseillé [76].

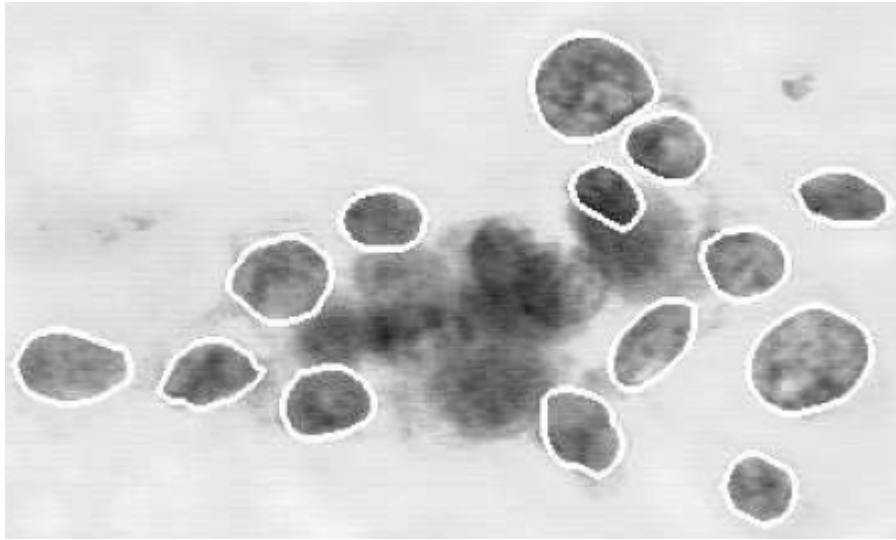
## 2.2 Caractéristique du noyau d'une cellule

Commençant par une limite approximative définie par l'utilisateur en tant qu'initialisation (figure A.1), la limite réelle du noyau de la cellule est localisée par un modèle de contour actif connu dans la littérature sous le nom de *'snake'*. Un *snake* est une spline déformable qui cherche à minimiser la fonction énergie définie sur la longueur d'une courbe fermée. La fonction énergie est définie de manière à ce que la valeur minimale apparaisse lorsque la courbe correspond exactement à la limite du noyau d'une cellule [76].

FIGURE A.1 – Limites initiales approximatives des noyaux de cellules



Afin de contrôler le temps de calcul, la valeur locale optimale de la fonction énergie est approximée à l'aide d'un algorithme glouton. Si la valeur de la fonction à un point de *snake* particulier peut être abaissée en déplaçant le point sur un pixel adjacent, il est déplacé, ce qui a éventuellement un effet sur le calcul de l'énergie en d'autres points. Le processus est répété pour chaque point jusqu'à ce que tous les points se fixent dans un minimum local de la fonction d'énergie. Les résultats d'une image typique sont illustrés à la figure A.2 .

FIGURE A.2 – *snakes* après la convergence vers les noyaux cellulaires

Le système de diagnostic qui se base sur la vision par ordinateur extrait dix caractéristiques différentes avec limites des noyaux cellulaires. Toutes les caractéristiques sont modélisées numériquement de manière à augmenter les risques de malignité.

Toutes les caractéristiques de forme ont été vérifiées à l'aide de cellules fantômes idéalisées. Il a été montré que ces valeurs augmentent lorsque que les limites deviennent moins régulières et moins corrélées avec la taille du contour. La valeur moyenne, la plus grande valeur et l'erreur type de chaque caractéristique sont calculées pour chaque image.

Les plus grandes valeurs sont les plus intuitivement utiles pour le problème en question, car seules quelques cellules malignes peuvent apparaître dans un échantillon donné.

Les caractéristiques extraites à partir de ces cellules sont les suivantes :

**1. Rayon :** Le rayon d'un noyau individuel est mesuré par la longueur des segments de la ligne radiale à partir du centre et des points individuels du *snake*.

**2. Périmètre :** La distance totale entre les points du *snake* constitue le périmètre nucléaire.

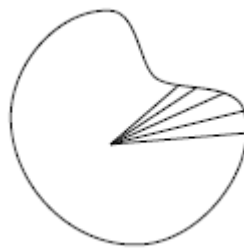
**3. Surface :** La surface nucléaire est mesurée simplement en comptant le nombre de pixels à l'intérieur du *snake* et en ajoutant la moitié des pixels dans le périmètre.

**4. Compacité :** Le périmètre et la surface sont combinés pour donner une mesure de la compacité des noyaux cellulaires en utilisant la formule du  $Perimetre^2/Surface$ . Ce nombre sans dimension est minimisé par un disque circulaire et augmente avec l'irrégularité de la limite. Cependant, cette mesure de forme augmente également pour les noyaux de cellules allongées, ce qui n'indique pas nécessairement une probabilité accrue de malignité.

La caractéristique est également biaisée vers le haut pour les petites cellules en raison de la précision réduite imposée par la numérisation de l'échantillon( voir figure A.3).

**5. Le lissage :** Le lissage d'un contour est quantifié en mesurant la différence entre la longueur d'une ligne radiale et la longueur moyenne des lignes l'entourant. Ceci est similaire au calcul de l'énergie de courbure chez les *snakes*. Voir la figure A.3.

FIGURE A.3 – Lignes radiales utilisées pour le calcul du lissage



**6. Concavité :** Dans une autre tentative pour capturer des informations de forme, le nombre et la gravité des concavités où des empreintes dans le noyau d'une cellule sont mesurés. Des accords entre des points du *snake* non adjacents sont établis et la limite réelle du noyau qui se trouve à l'intérieur de chaque accord est mesurée (voir la figure A.4).

Ce paramètre est grandement influencé par la longueur de ces accords, car les

FIGURE A.4 – Accords utilisés pour calculer la concavité

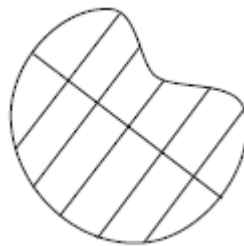


accords plus petits capturent mieux les petites concavités. Ils ont choisi de mettre l'accent sur les petites empreintes, car les irrégularités de forme plus grandes sont capturées par d'autres caractéristiques.

**7. Points concaves :** Cette caractéristique est similaire à la concavité mais ne mesure que le nombre et non la magnitude des concavités de contour.

**8. Symétrie :** Pour mesurer la symétrie, on trouve l'axe principal, où l'accord le plus long passe par le centre. Ensuite la différence de longueur entre les lignes perpendiculaires à l'axe principal de la limite de cellule dans les deux sens est mesurée. Voir Figure A.5.

FIGURE A.5 – Segments utilisés dans le calcul de symétrie



**9. Dimension fractale :** La dimension fractale d'une cellule est approximée à l'aide de l'approximation du trait de côte "décrite par Mandelbrot [77].

Le périmètre du noyau est mesuré à l'aide de «règles» de plus en plus grandes. À mesure que la taille de la règle augmente, la précision et le périmètre observé diminuent. Voir la figure A.6. La représentation graphique de ces valeurs sur une échelle logarithmique et la mesure de la pente descendante donnent une approximation de la dimension fractale. Comme pour toutes les caractéristiques de

forme, une valeur plus élevée correspond à un contour moins régulier et donc à une probabilité plus élevée de malignité.

FIGURE A.6 – Séquence de mesures pour calculer la dimension fractale



**10. Texture :** La texture du noyau de la cellule est mesurée en fonction de la variance des intensités de gris dans les pixels composants.

Toutes les caractéristiques de forme ont été vérifiées à l'aide de cellules fantômes idéalisées. Il a été démontré qu'ils augmentaient à mesure que les limites devenaient moins régulières.

Il faut noter que, la valeur moyenne, la valeur extrême (la plus grande) et l'erreur standard de chaque caractéristique sont calculées pour chaque cellule.

### 3 le Cardiotographie

Parmi les désirs communs de toutes les futures mères est le bien être de leur bébé [83].

Les obstétriciens déterminent le mode d'accouchement approprié en cas de situation anormale en fonction du résultat d'un diagnostic, le risque ou l'évolution défavorable du fœtus pouvant alors être minimisé et même éliminé [80]. Par conséquent, la détection et le diagnostic de l'état de la mère et du fœtus sont essentiels et nécessaires pour la mère et le bébé pendant la grossesse.

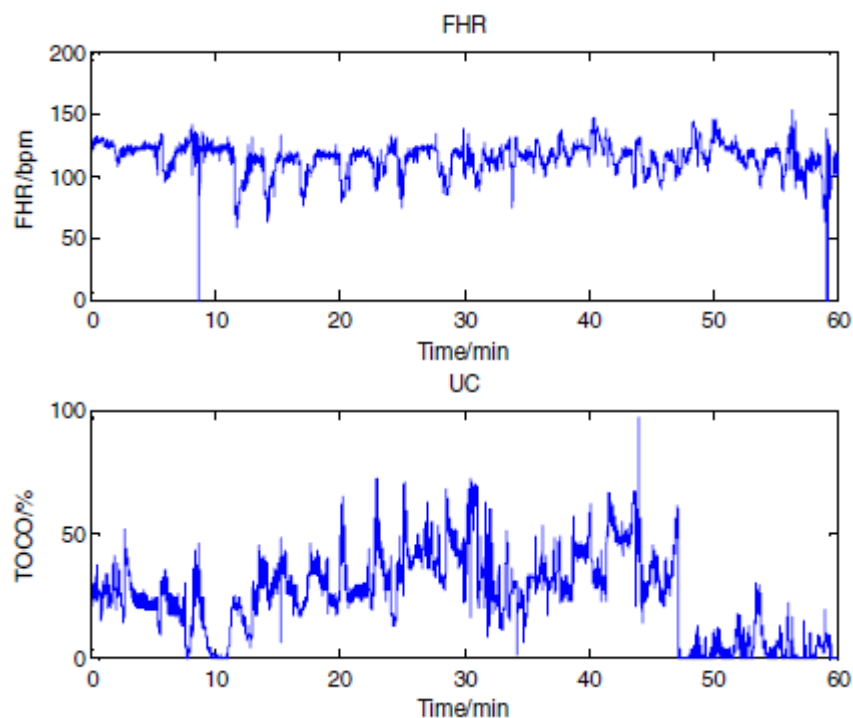
Pour atteindre l'objectif ci-dessus, la cardiotocographie (CTG) est un instrument non invasif et efficace pour surveiller le statut foetal et évaluer le bien-être du fœ-

tus lors d'un examen prénatal pendant les périodes ante-partum.

Le CTG est connu également sous le nom de surveillance fœtale électronique (EFM), comprend : la surveillance simultanée du signal de la fréquence cardiaque fœtale (FHR), via une sonde Doppler à ultrasons, et des Contractions Utérines (UC), au moyen d'un transducteur de pression, tous deux appliqués à l'extérieur de l'abdomen de la femme enceinte . Cette méthode est meilleure que d'autres méthodes antérieures, telles que le prélèvement de sang sur le cuir chevelu fœtal et l'auscultation périodique [81].

La figure A.7 illustre des échantillons numérisés d'un signal CTG typique obtenu à partir de CTUUHB, une base de données à accès ouvert pour cardiocardiographie intrapartum. [82]

FIGURE A.7 – Un enregistrement CTG typique comprenant FHR et UC.



Le cardiocardiogramme (CTG) est le moyen le plus largement utilisé pour évaluer l'état du fœtus. La machine utilisée pour effectuer la surveillance s'appelle un cardiocardiographe, plus communément appelé moniteur fœtal électronique (EFM).

La surveillance par CTG est largement utilisée pour évaluer le bien-être du fœtus. [79] Une étude a révélé que les machines CTG informatisées entraînaient un nombre de décès de bébés inférieur à celui des machines CTG traditionnelles.

Le CTG consiste en deux traces, l'une représentant la fréquence cardiaque fœtale (FHR) et l'autre, l'activité des UCs. De nombreuses méthodes automatiques ont été proposées pour l'interprétation du CTG. La plupart d'entre eux s'appuient soit sur une approche de classification binaire, soit sur une approche multi-classes.

L'apprentissage automatique a pour objectif de classer l'état du fœtus, comme pathologique ou normal en se basant sur les attributs extraits des signaux FHR et UC, ce qui aide les obstétriciens à prendre des meilleures décisions.

### 3.1 Description de la base de Données CTG

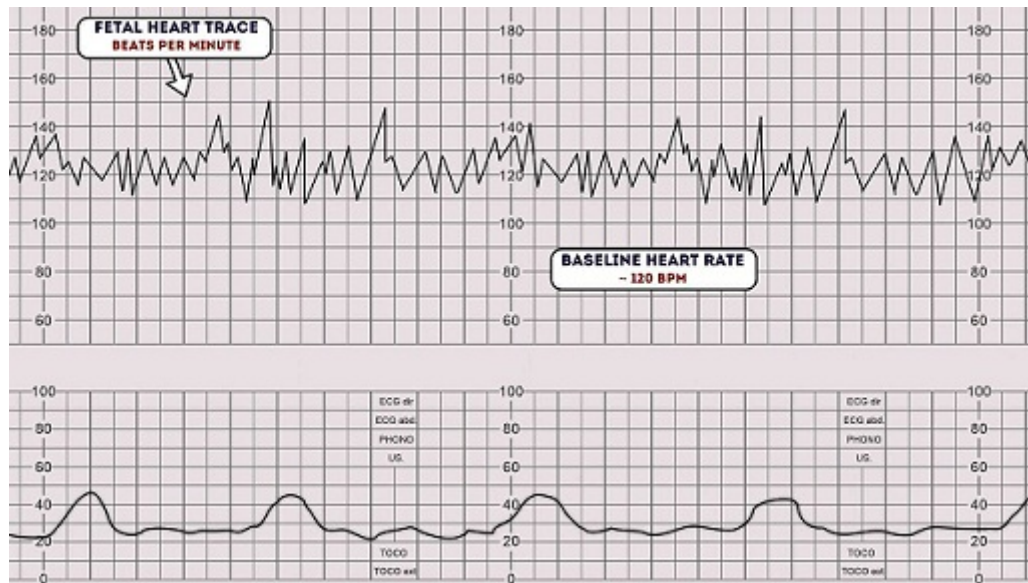
#### 1. LB : Valeur de base de la fréquence cardiaque fœtale (Fetal Heart Rate Baseline Value)(FHR)

La valeur de base de la FHR est définie comme la fréquence cardiaque approximative entre les contractions des utérines (voir figure A.8). La plage normale est comprise entre 110 et 160 bpm. Les taux de référence inférieurs à cette plage (moins de 110 bpm) sont considérés comme de la bradycardie, alors que la tachycardie de référence fœtale est définie par une FCF supérieure à 160 bpm pour deux cycles de contraction supérieure à 5 minutes. Les écarts par rapport à cette plage de base normale ne sont pas nécessairement indicatifs d'un compromis fœtal, s'il n'est pas associé à d'autres anomalies de la FHR .

#### 2. AC : Les accélérations de la FHR

Les accélérations périodiques de la FHR intrapartum (plus de 14 battements par minute au-dessus de la valeur initiale durant plus de 14 secondes et moins de 2 minutes) 14 peuvent être spontanées, associées à des contractions utérines ou liées aux mouvements du fœtus. Comme dans le cadre de la surveillance prénatale (par exemple, test sans stress), les accélérations intrapartum de la FCF fournissent des preuves du fonctionnement intact du système nerveux central fœtal

FIGURE A.8 – Valeur de base de la fréquence cardiaque fœtale



et, donc, du bien-être fœtal. La présence d'accélération de la FHR spontanées est clairement un signe rassurant. L'inverse, cependant, n'est pas vrai ; l'absence de telles accélérations n'est pas nécessairement inquiétante et d'autres caractéristiques de traçage de la FHR devraient être utilisées pour évaluer le statut fœtal.

### 3. MF : Les mouvements du fœtus :

La plupart des futures mères savent comment leurs bébés bougent et interprètent intuitivement le ralentissement des mouvements du fœtus (RFM) comme un signe alarmant. L'inquiétude maternelle pour la RFM est une cause fréquente de consultations prénatales non programmées, qui se produisant dans environ 10% des grossesses du troisième trimestre, bien que les directives existantes recommandent que les femmes enceintes soient informées de la nécessité de contacter un professionnel de la santé qui leur donne des informations ambiguës sur ce à quoi s'attendre concernant l'activité fœtale.

Selon une opinion largement répandue, le RFM est normal à l'approche de la grossesse. Cependant, les preuves à l'appui sont limitées et des incertitudes subsistent quant à la définition des modèles de MF normaux [84]. De plus, on ignore ce qui constitue un changement cliniquement important dans la perception maternelle de la FM, ce qui compromet le comptage de la FM comme un moyen d'identifier de manière prospective les grossesses présentant un risque d'évolution défavorable. À l'heure actuelle, la définition la mieux fondée du RFM pro-

vient de l'étude de Moore et Piacquadio, qui définissait la RFM comme la perception maternelle de moins de dix mouvements en moins de deux heures (ci-après dénommé «l'alarme de deux heures») [85]. Soit environ cinq écarts-types par rapport au temps de comptage moyen lors de grossesses normales. Les mouvements du fœtus extraits à partir du tractage de la FHR sont en nombre de mouvement par seconde.

#### 4. Les contractions utérines :

Les contractions utérines sont quantifiées comme le nombre de contractions présentes dans une durée de 10 minutes, moyennées sur 30 minutes [87]. La fréquence de contraction seule est une évaluation partielle de l'activité utérine. D'autres facteurs tels que la durée, l'intensité et le temps de relaxation entre les contractions sont également importants en pratique clinique.(voir figure A.9)

Ce qui suit représente la terminologie pour décrire l'activité utérine :

**A. Normal :** moins de 5 contractions en 10 minutes en moyenne sur une durée de 30 minutes.

**B. Tachysystole :** plus que 5 contractions en 10 minutes en moyenne sur une durée de 30 minutes.

#### C. Caractéristiques des contractions utérines :

- La tachysystole doit toujours être qualifiée quant à la présence ou à l'absence de décélérations associées à la FCF.
- Le terme tachysystole s'applique tant au travail spontané qu'au travail stimulé. La réponse clinique au tachysystole peut varier selon que les contractions sont spontanées ou stimulées.
- Les termes hyperstimulation et hypercontractilité ne sont pas définies et doivent être abandonnés.

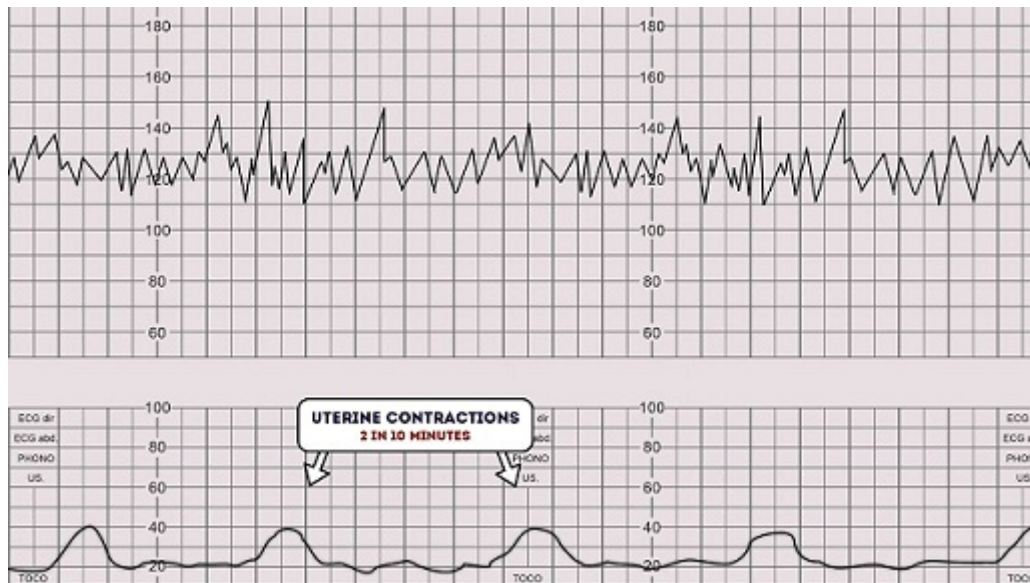
Dans la base de données CTG, ces contractions sont enregistrées en nombre de contractions par seconde.

#### 5. Décélération tardive :

Les décélérations tardives commencent au plus fort de la contraction utérine et se rétablissent une fois la contraction est terminée (voir figure A.10).

Ce type de décélération indique que le flux sanguin vers l'utérus et le placenta est insuffisant [87]. En conséquence, le flux sanguin vers le fœtus est considérablement réduit, entraînant une hypoxie et une acidose fœtales.

FIGURE A.9 – Les contractions utérines



### 6. Décélération variable :

Les décélérations variables sont observées comme une chute rapide du rythme cardiaque fœtal de base avec une phase de récupération variable (voir figure A.11). Ils sont de durée variable et peuvent ne pas avoir de relation avec les contractions utérines [87]. Ils sont le plus souvent observés pendant le travail et chez les patients présentant un volume de liquide amniotique réduit.

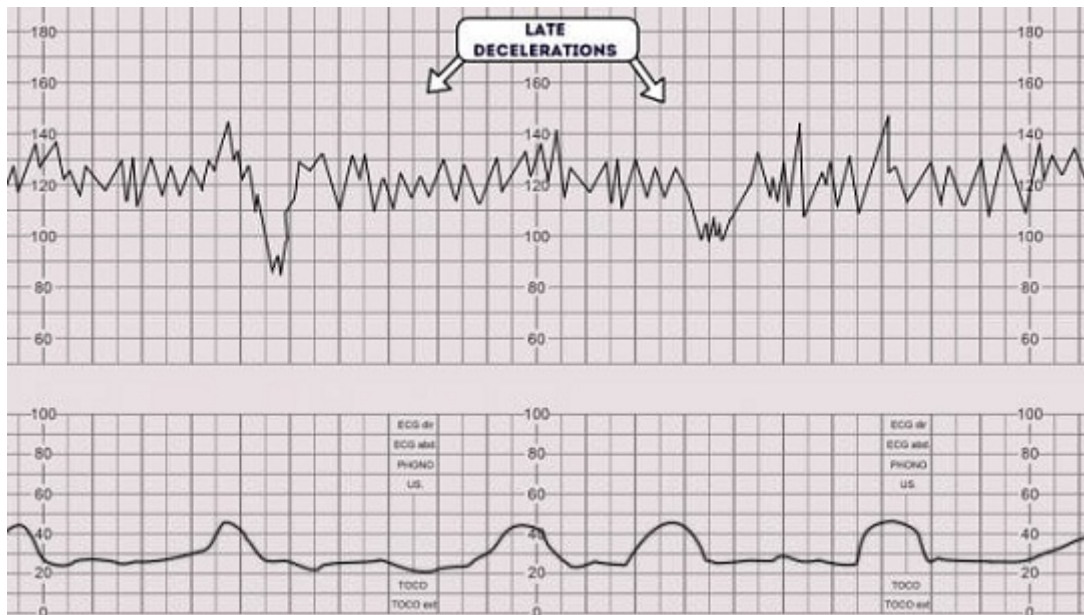
### 7. Décélération prolongée :

Une décélération prolongée est définie comme une décélération qui dure plus de 3 minutes (voir figure A.12). Si c'est le cas, elle est immédiatement classée comme anormale [87].

### 8. Variation de la fréquence cardiaque fœtale à court terme :

La relation entre la réduction de la variation du rythme cardiaque fœtal et l'acidémie fœtale a été analysée à l'aide d'un système automatique [88,89]. Entre 1983 et 1987, ils ont identifié 78 grossesses dans lesquelles au moins un enregistrement de la fréquence cardiaque fœtale présentait une très faible variation à long terme. Le résultat a été analysé afin de déterminer les critères numériques de variation de la fréquence cardiaque fœtale permettant de détecter le plus efficacement possible l'état du fœtus. Parce que la compromission fœtale a parfois été associée à un rythme cardiaque sinusoïdal lent, augmentant les mesures de la variation à long terme. Il était nécessaire de définir un nouvel indice de variation de la fréquence cardiaque fœtale à court terme (variation de 6 minutes). Ceci était étroitement lié

FIGURE A.10 – La décélération tardive



aux variations à long terme, mais permettait une meilleure détection des enregistrements prématurés, à en juger par l'acidémie métabolique à la naissance ou le décès intra-utérin A.13.

### 9. MSTV

La valeur moyenne de la variabilité à court terme.

### 10. Variation de la fréquence cardiaque fœtale à long terme :

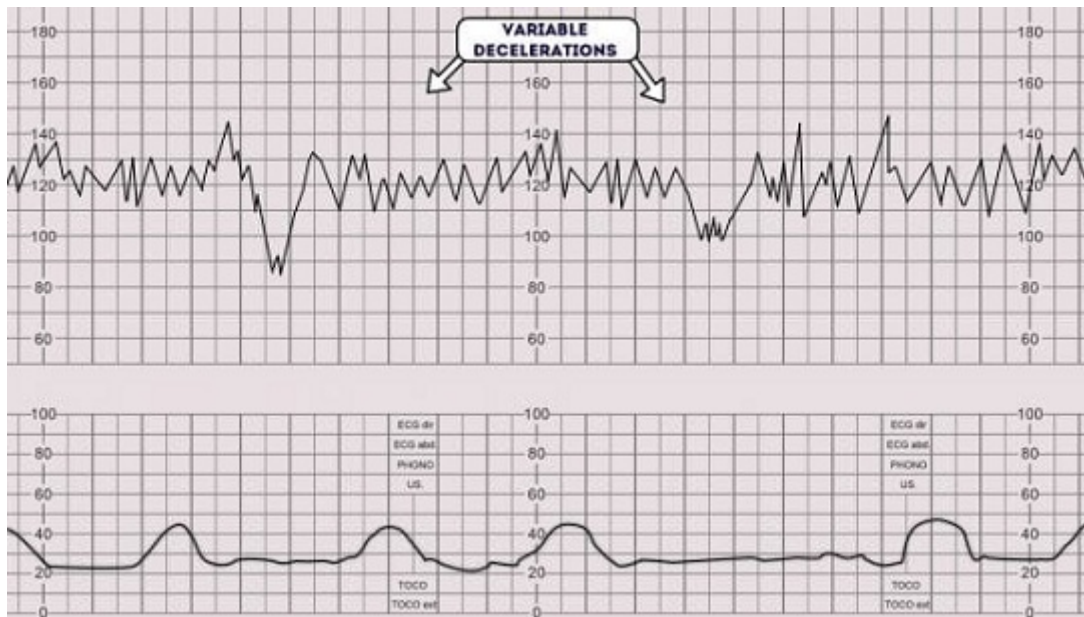
La variabilité à long terme représente des oscillations générales de la fréquence cardiaque fœtale, qui se produisent plusieurs fois par minute. L'accélération du cœur fœtal est une forme de variabilité à long terme [89]. Celles-ci surviennent généralement en réponse aux mouvements du fœtus et dépassent de 15 BPM la valeur de base, pendant 10 à 20 secondes ou plus. Ils peuvent souvent être provoqués en stimulant le fœtus lors d'un examen pelvien ou en stimulant acoustiquement le fœtus avec un bruit fort. La présence d'accélération fœtales rassure sur le fait que le fœtus est en bonne santé et tolère bien l'environnement intra-utérin A.14. .

### 11. MLTV

La valeur moyenne de la variabilité à long terme.

**Les histogrammes :** Les histogrammes et les histogrammes normalisés des

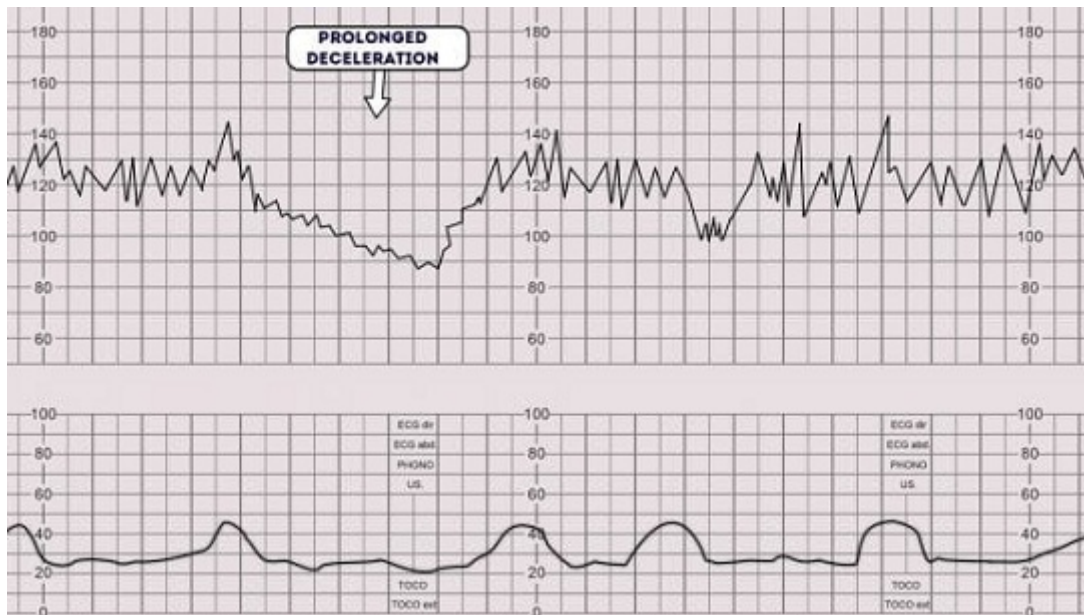
FIGURE A.11 – La décélération variable :



variables prédictives sont aussi analysés pour comprendre la distribution des données et la relation entre les variables prédictives et cibles, parmi les attributs tirés à partir de ces histogrammes :

12. **Largeur** - largeur de l'histogramme FHR
13. **Min** - minimum de l'histogramme FHR
14. **Max** - maximum de l'histogramme FHR
15. **Nmax** - Maximum de pics d'histogramme
16. **Nzeros** - Maximum de zéros d'histogramme
17. **Mode** - mode histogramme
18. **Moyenne** - histogramme moyen
19. **Médiane** - médiane de l'histogramme

FIGURE A.12 – La décélération prolongée :



20. **Variance** - Variance de l'histogramme

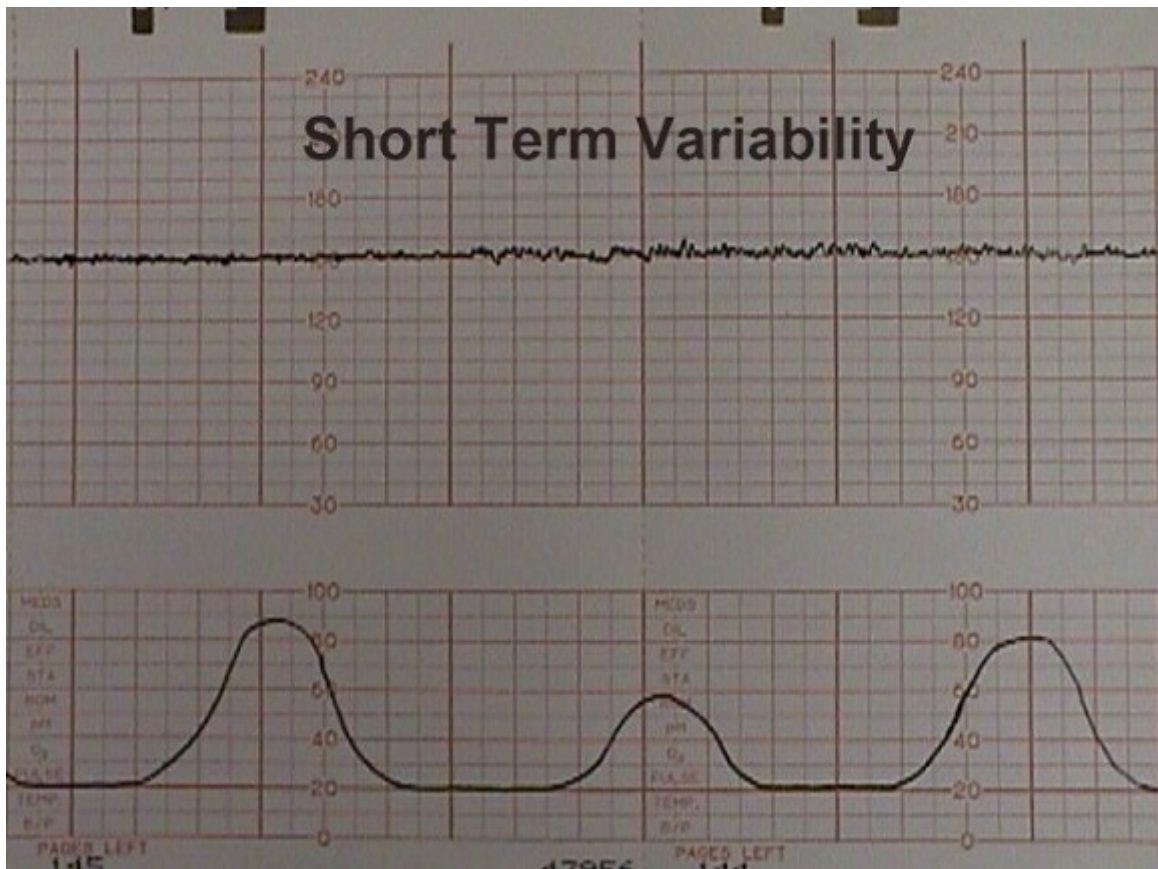
21. **Tendance** - tendance de l'histogramme

22. **CLASS** - Code de classe de modèle FHR (1 à 10)

Les classes sont détaillées avec ces abréviations [90] :

- **A** : Sommeil calme
- **B** : Sommeil (mouvement oculaire rapide)
- **C** : Vigilance calme
- **D** : Vigilance active
- **SH** : Modèle de changement (A ou SUSP avec changements)
- **AD** : Modèle d'accélération / décélération (situation de stress)

FIGURE A.13 – La variation de la fréquence cardiaque fœtale à court terme :

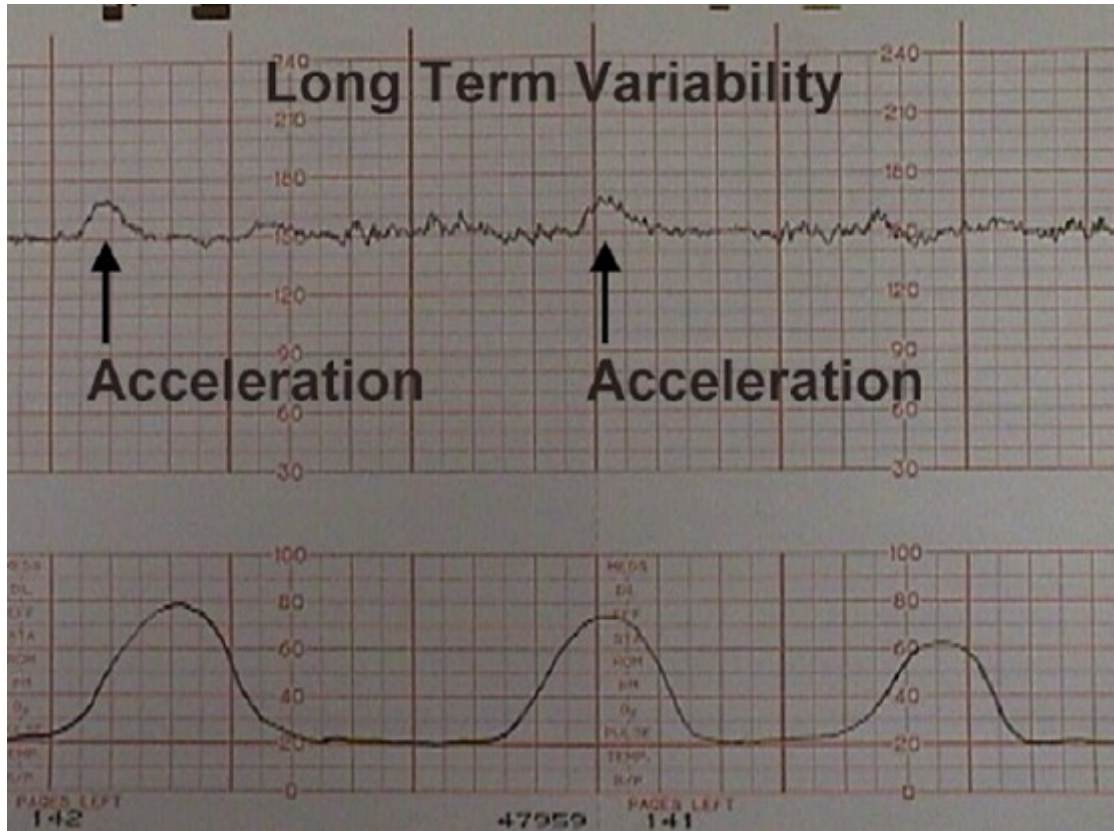


- **DE** : Schéma de décélération (stimulation vagale)
- **LD** : Largement décélératif
- **FS** : Modèle plat et sinusoidal (état pathologique)
- **SUSP** : Modèle de suspect

## 22. NSP - code de classe d'état fœtal

- **N** : Normal
- **S** : Suspect

FIGURE A.14 – La variation de la fréquence cardiaque fœtale à court terme :



– P : Pathologique

## 4 Hépatite

L'hépatite gonfle le foie, l'empêchant de bien fonctionner. Cela peut entraîner un cancer. Les virus causent la plupart des cas d'hépatite. Les différents types de la maladie sont nommés pour les virus qui les causent. Par exemple, le virus de l'hépatite A est la cause de l'hépatite A. La consommation de drogue ou d'alcool peut également entraîner une hépatite. Dans d'autres cas, le système immunitaire attaque par erreur les cellules hépatiques saines du corps. Les vaccins préviennent certaines formes d'hépatite virale. Parfois, l'hépatite disparaît d'elle-même. Si ce n'est pas le cas, il peut être traité avec des médicaments.

### Types d'hépatite :

Actuellement, cinq virus différents sont connus pour causer l'hépatite :

- **Hépatite A** : Parfois appelée «hépatite infectieuse», l'hépatite A se transmet par la consommation d'aliments ou d'eau potable contaminée par des déchets humains. L'hépatite A met rarement la vie en danger.
- **Hépatite B** : Également appelée «hépatite sérique», l'hépatite B se transmet de la mère à l'enfant à la naissance ou peu après, ainsi que par contact sexuel, par transfusion de sang et par aiguille contaminés. L'hépatite B peut provoquer une cicatrice du foie (cirrhose) et entraîner un cancer du foie.
- **Hépatite C** : Anciennement appelée «hépatite non-A, non-B», l'hépatite C est la forme la plus courante d'hépatite virale. Bien qu'il puisse être transmis par des transfusions de sang et/ou des aiguilles contaminées, la cause n'est pas connue pour un grand nombre de patients. Cela pourrait marquer le foie. L'infection à l'hépatite C est courante chez environ 25% des personnes séropositives. L'hépatite C infecte également jusqu'à 90% des consommateurs de drogues injectables infectés par le VIH. Et il est plus grave chez les patients infectés par le VIH.
- **Hépatite D** : Cette forme infecte le plus souvent les utilisateurs de drogues par voie intraveineuse (IV) qui sont également porteurs du virus de l'hépatite B. Il ne se propage qu'en présence du virus de l'hépatite B et se transmet de la même manière. L'hépatite D est un grave problème de santé car elle survient chez les personnes atteintes d'hépatite B, ce qui accroît la gravité des symptômes associés à l'hépatite B.
- **Hépatite E** : Comme l'hépatite A, l'hépatite E est répandue dans les pays où l'assainissement est médiocre. Il est rare en Amérique du Nord et met rarement la vie en danger.

#### **Hépatite aiguë :**

L'hépatite aiguë est l'infection initiale, qui peut être légère ou grave. Si l'infection dure six mois ou plus, on parle alors d'hépatite chronique. Les hépatites A et E ne provoquent pas d'hépatite chronique. Les virus de l'hépatite B, C et D peuvent provoquer des épisodes aigus et chroniques. Les hépatites chroniques B et C sont particulièrement graves.

#### **Description des attributs de la base de Données Hépatite :**

1. **Classe** : Mort, Vivant
2. **AGE** : 10, 20, 30, 40, 50, 60, 70, 80
3. **SEXE** : homme, femme
4. **STEROID** : non, oui  
 Les stéroïdes forment une vaste famille chimique qui regroupe toutes les molécules naturelles dont le squelette carboné polycyclique s'apparente au gonane. Ils sont très répandus dans la nature où on les rencontre à tous les échelons du règne végétal et du règne animal.
5. **ANTIVIRAUX** : non, oui  
 Un antiviral est une molécule gênant le cycle de réplication d'un ou de plusieurs virus, permettant ainsi de retarder et d'arrêter une infection virale.
6. **FATIGUE** : non, oui
7. **MALAISE** : non, oui
8. **ANOREXIE** :(manque d'appétit) : non, oui
9. **GRAND FOIE** : non, oui
10. **FERME DU FOIE** : non, oui  
 C'est est une couleur jaune de la peau, des muqueuses ou des yeux provoquée par une accumulation de bilirubine dans le sang. La jaunisse est classée en trois catégories : préhépatique, hépatocellulaire et posthépatique.
11. **Rate palpable** : non, oui
12. **Araignées** : non, oui
13. **ASCITES** : non, oui  
 L'ascite est la présence d'un excès de liquide dans la cavité péritonéale
14. **VARICES** : non, oui
15. **BILIRUBINE** : 0,39, 0,80, 1,20, 2,00, 3,00, 4,00  
 La bilirubine est habituellement normale ou peu augmentées. Dans le cas contraire, un syndrome de chevauchement avec la la cirrhose biliaire primitive ou la cholangite sclérosante primaire doit être recherché. Une élévation des gammaglobulines est retrouvée dans 90
16. **PHOSPHATE D'ALK** : 33, 80, 120, 160, 200, 250  
 La mesure de la phosphatase alcaline permet aussi de mettre en évidence une atteinte hépatobiliaire.

17. **SGOT** : 13, 100, 200, 300, 400, 500,

Le test SGOT est un test sanguin qui fait partie d'un profil hépatique. Il mesure l'une des deux enzymes hépatiques, appelée transaminase sérique glutamique-oxaloacétique.

18. **ALBUMINE** : 2.1, 3.0, 3.8, 4.5, 5.0, 6.0

Les modifications protéiques sériques survenant dans les maladies du foie associées aux lésions parenchymateuses sont typiquement réduites dans l'album sérique.

19. **PROTIME** : 10, 20, 30, 40, 50, 60, 70, 80, 90

Le protime est un test utilisé pour évaluer la coagulation du sang. Les facteurs de coagulation sanguine sont des protéines fabriquées par le foie. Lorsque le foie est gravement atteint, ces protéines ne sont normalement pas produites. Le taux de prothrombine est également un test utile de la fonction hépatique, car il existe une bonne corrélation entre les anomalies de la coagulation mesurées par le temps de prothrombine et le degré de dysfonctionnement hépatique.

20. **HISTOLOGIE** : non, oui

L'histologie hépatique montre simultanément les caractéristiques des hépatites chroniques.