

Modélisation en Aide à la Décision

Support de Cours Présenté Par

Dr. SOUICI Ismahane

Maître de Conférences B - Département d'Informatique
Faculté des Sciences Exactes et Informatique
Université de Jijel

Avril 2016

Ce support du cours de modélisation en aide à la décision est destiné aux
étudiants de Master 1

Option : SIAD, Systèmes d'Informations et Aide à la Décision

Département d'Informatique

Faculté de sciences exactes et informatique

Université de Jijel

Table des matières

| | |
|--------------------------|----|
| Liste des figures | I |
| Liste des tableaux | II |

| | |
|----------------------------|---|
| Introduction générale..... | 8 |
|----------------------------|---|

Chapitre I : INTRODUCTION A L'AIDE A LA DECISION

| | |
|---|----|
| 1. Introduction | 9 |
| 2. Décision | 10 |
| 2.1. La notion de la décision..... | 10 |
| 2.2. Décideur | 11 |
| 2.3. Types de décisions | 11 |
| 3. Aide à la décision | 12 |
| 4. Le décisionnel | 13 |
| 5. Le processus de décision | 13 |
| 6. Historique des systèmes décisionnels | 15 |
| 7. Références bibliographiques..... | 16 |

Chapitre II : SYSTEMES A BASE DE CONNAISSANCES

| | |
|--|----|
| 1. Introduction | 18 |
| 2. La notion d'un système à base de connaissances | 18 |
| 2.1. La résolution de problèmes..... | 18 |
| 2.2. Connaître c'est décrire pour retrouver | 19 |
| 3. L'architecture et la mécanique des systèmes de connaissances..... | 19 |
| 3.1. Les concepts et les ontologies | 19 |
| 3.2. L'ingénierie et la gestion de connaissances..... | 20 |
| 4. Les formalismes logiques et Prolog | 21 |
| 4.1. Le calcul des propositions | 21 |
| 4.2. Le calcul des prédicats du premier ordre | 21 |
| 4.3. Prolog..... | 22 |
| 5. Les systèmes à base de règles | 22 |
| 6. Histoire, applications et enjeux actuels | 23 |
| 7. Références bibliographiques | 23 |

Chapitre III : LES SYSTEMES DECISIONNELS

| | |
|---|----|
| 1. Le besoin d'un système décisionnel | 25 |
| 1.1. Le système d'information | 25 |
| 1.2. Système d'information décisionnel..... | 25 |
| 2. Outils d'aide à la décision | 25 |
| 2.1. Typologie des outils d'aide à la décision | 25 |
| 2.1.1. Classement en fonction du niveau de décision..... | 25 |
| 2.1.2 Classement en fonction de l'envergure de la décision | 26 |
| 2.2. Les qualités indispensables d'un outil d'aide à la décision..... | 27 |
| 3. Les composants de base d'un système décisionnel..... | 27 |
| 4. Les Systèmes Interactifs d'Aide à la Décision (SIAD)..... | 27 |
| 4.1. Définition des SIAD | 28 |

| | |
|--|----|
| 4.2. Caractéristiques des SIAD | 28 |
| 4.3. Composition d'un SIAD | 29 |
| 4.4. Types d'architecture des SIAD | 30 |
| 5. Conclusion | 30 |
| 6. Références bibliographiques | 30 |

Chapitre IV : LES ARBRES DE DECISION

| | |
|---|----|
| 1. Introduction | 32 |
| 2. Représentation par arbre de décision..... | 32 |
| 3. Objectif | 33 |
| 4. Construction d'un arbre de décision | 33 |
| 5. Apprentissage d'un arbre de décision | 36 |
| 5.1. Choix d'une variable de segmentation..... | 36 |
| 5.2. Traitement des variables continues..... | 38 |
| a. Choix du point de coupure..... | 38 |
| b. Sélectionner la variable de segmentation | 39 |
| 5.3. Définir la bonne taille de l'arbre | 40 |
| a. Pré-élagage | 40 |
| b. Post-élagage | 41 |
| 5.4. Décision | 42 |
| 5.5. Fusion des sommets lors de la segmentation..... | 42 |
| 6. Avantages et faiblesses des arbres de décision | 44 |
| 6.1. Avantages | 44 |
| 6.2. Faiblesses | 44 |
| 7. Evaluation d'un arbre de décision | 45 |
| 7.1. Les critères de décision non probabilistes | 45 |
| 7.2. Décision sous risque | 45 |
| 8. Références bibliographiques | 46 |

Chapitre V : LES DIAGRAMMES D'INFLUENCE

| | |
|--|----|
| 1. Introduction | 49 |
| 2. Composants d'un DI | 49 |
| 2.1. Les variables | 49 |
| 2.2. Les formules | 49 |
| 2.3. Les flèches (arcs) | 49 |
| 3. Conception d'un diagramme d'influence | 51 |
| 4. Caractéristiques des diagrammes d'influence | 52 |
| 5. Références bibliographiques | 52 |

Chapitre VI : LES RESEAUX BAYESIENS

| | |
|---|----|
| 1. Introduction | 54 |
| 2. Probabilités | 54 |
| 2.1. Définitions principales | 55 |
| 2.2. Probabilités sur plusieurs variables..... | 56 |
| 3. Les réseaux bayésiens | 58 |
| 4. Graphe de causalité (Représentation qualitative) | 58 |
| 4.1. Représentation de la causalité | 58 |
| 4.2. Circulation de l'information dans un graphe de causalité | 59 |
| 4.2.1. Exemple | 59 |

| | |
|--|----|
| 4.2.2. Indépendance conditionnelle et d-séparation | 60 |
| 5. Représentation probabiliste associée (quantitative) | 62 |
| 5.1. Transposition qualitative / quantitative | 62 |
| 5.1.1. Définition des variables | 63 |
| 5.1.2. Définition des probabilités | 63 |
| 5.2. Etude de cas № 01 | 64 |
| 5.2.1. Formalisation qualitative / quantitative | 64 |
| 5.2.2. Discussion | 65 |
| 5.3. Etude de cas № 02 | 66 |
| 5.3.1. Formalisation qualitative / quantitative | 66 |
| 5.4. Etude de cas № 03 | 69 |
| 6. Références bibliographiques | 70 |
| Conclusion générale..... | 72 |
| Annexe 1 : Exercices | 73 |
| Annexe 2 : Examens 2009 - 2015 | 78 |

Listes des figures

| | |
|---|----|
| Figure I.1. Lien entre les mesures, la stratégie et la décision | 10 |
| Figure I.2. La structuration intervient avant la normalisation | 12 |
| Figure I.3. Procédé d'extraction des connaissances | 13 |
| Figure I.4. Le modèle I.D.C | 14 |
| Figure III.1. Principe des SIAD d'après Bonczek | 28 |
| Figure III.2. Principe des SIAD suivant Spague | 28 |
| Figure III.3. Composition d'un SIAD | 29 |
| Figure IV.1. Arbre de décision modélisant la base d'exemples 1 | 33 |
| Figure IV.2. Arbre de décision modélisant la base de données "weather" | 35 |
| Figure IV.3. Sélection de la borne de discrétisation | 38 |
| Figure IV.4. Variation du taux d'erreur en apprentissage et en test | 40 |
| Figure IV.5. Segmentation de la racine de l'arbre avec fusion des sommets enfants | 44 |
| Figure V.1. Interprétations de relations | 50 |
| Figure V.2. Exemple du parapluie en diagramme d'influence | 51 |
| Figure VI.1. Graphe de causalités modélisant l'exemple de l'arroseur de Pearl | 59 |
| Figure VI.2. Types de connexions entre trois nœuds, (a) en série, (b) divergente, (c) convergente | 60 |
| Figure VI.3. Equivalence entre représentation qualitative et quantitative | 66 |

Listes des tableaux

| | |
|---|----|
| Table IV.1. Base d'exemples 1 | 33 |
| Table IV.2. Données "weather" | 34 |
| Table IV.3. Tri croisé à l'aide de la variable "ensoleillement" à la racine de l'arbre | 36 |
| Table IV.4. Tableau des effectifs lors du croisement de deux variables | 36 |
| Table IV.5. Descripteurs discrets candidats sur la racine de l'arbre | 37 |
| Table IV.6. Segmentation candidates et bornes de discrétisation associées pour les descripteurs continus | 39 |
| Table IV.7. Règles extraites de l'arbre de la Figure IV.2 | 42 |
| Table IV.8. 1ère passe, fusion deux à deux des sommets pour la segmentation de la racine "ensoleillement" | 43 |
| Table IV.9. 2ème passe, fusion des sommets pour la segmentation de la racine "ensoleillement" | 44 |
| Table IV.10. Base d'exemples : Investissement immobilier | 45 |
| Table VI.1. Distribution des événements élémentaires en fonction d'une valeur aléatoire | 55 |
| Table VI.2. Exemples illustratifs de la circulation de l'information suivant les types de connexions de nœuds | 61 |

Introduction générale

Aujourd'hui, les données brutes utilisées dans les diverses activités de l'entreprise (données clients, fournisseurs, statistiques et observation de la production, etc.) sont considérées comme un actif réel de l'entreprise. Ce qui fait que la prise de décision nécessite de prendre en compte une quantité croissante de données, d'informations et de connaissances de différentes natures et qualités. La compétitivité passe par leur analyse et leur optimale exploitation. Ainsi, ce qu'on appelle le décisionnel (extraire depuis les données brutes des informations et des indicateurs pertinents pour la stratégie de l'entreprise) prend toute sa place dans l'entreprise.

En conséquence, les gestionnaires du risque ont de plus en plus recours à l'informatique pour se doter d'outils puissants d'aide à la décision. Ce support de cours a pour objectif d'apporter les connaissances nécessaires au développement de tels outils. Il est structuré en deux grandes parties, une première faisant le tour des notions et fondements théoriques nécessaires à l'exploration de l'informatique décisionnelle ; et une deuxième partie consacrée à la présentation de méthodes de modélisation en aide à la décision proprement dite. Deux annexes sont finalement proposés à la fin de ce support dans le but d'affermir la compréhension du cours. Le premier rassemble des exercices de natures différentes accompagnés de leurs corrigés types ou éléments de réponses. Le deuxième regroupe les examens ordinaires et les examens de rattrapage proposés depuis la première jusqu'à la sixième année d'enseignement du module.

La première partie recouvre les trois premiers chapitres dont le premier introduit le concept de décision, de décideur et une récapitulation des étapes du processus décisionnel. Autrement dit, il présente la problématique générale de l'aide à la décision, en privilégiant le contexte organisationnel. Le deuxième chapitre expose l'historique et le principe des systèmes à base de connaissances. Le troisième chapitre rappelle le besoin d'un système décisionnel, suivi de la présentation de la typologie des outils d'aide à la décision en résumant les qualités indispensables d'un outil d'aide à la décision et en proposant deux classements possibles de ces outils en fonction du niveau de décision ou de l'envergure de la décision. Ensuite, l'accent sera mis sur un type particulier d'outils d'aide à la décision à savoir les Systèmes Interactifs d'Aide à la Décision (SIAD), en présentant leurs définition, caractéristiques et composition.

La deuxième partie traite de la modélisation en aide à la décision. Elle regroupe l'ensemble des trois restants chapitres présentant chacun le détail nécessaire à la bonne compréhension et exploitation des méthodes de modélisation suivantes : arbres de décision, diagrammes d'influences et réseaux bayésiens utilisés pour des fins d'aide à la décision, bien sûr.

A la fin du contenu de chacun des six chapitres, l'étudiant cherchant plus de détails, trouvera des références importantes proposées.

Enfin, il sera utile de signaler que le contenu du cours a été amélioré pendant six années d'enseignement consécutives du module de "Modélisation en aide à la décision" depuis l'année universitaire 2009-2010 jusqu'à l'année universitaire 2014-2015. C'est pourquoi, ce document peut être d'usage important pour le développement des projets de fin d'études des étudiants en deuxième année de master SIAD.

Chapitre I : Introduction à l'aide à la décision

1. Introduction

Dans la vie quotidienne, nos décisions sont souvent prises sur la base d'intuitions et d'expériences passées. Elles sont issues d'heuristiques observables au travers de biais systématiques (Kahnemann et al., 1982). Mais en réalité ce type de stratégies ne peut s'appliquer qu'à des problèmes familiers. Lorsque nous sommes confrontés à des situations nouvelles, la tâche de prise de décision devient beaucoup plus difficile. De nos jours, l'environnement des décideurs est de plus en plus complexe et évolue rapidement. La tendance est plutôt à l'accroissement de cette complexité. Cet accroissement est dû à plusieurs facteurs :

- La technologie de l'information et des ordinateurs,
- une complexité structurelle des décisions,
- une plus forte compétition,
- l'ouverture au marché international,
- etc.

Ainsi, pour acquérir de nouvelles parts de marché, l'entreprise doit mieux comprendre ses clients et être réactive pour identifier les nouveaux relais de croissance. Dès lors, l'entreprise doit davantage prendre en compte l'évolution du marché afin d'en détecter les opportunités et les menaces. C'est ce que Le moigne (1974) a confirmé en annonçant que, la prise et l'exécution des décisions sont les buts fondamentaux de toute organisation, de tout management; et que toute organisation dépend, structurellement, de la nature des décisions qui sont prises en son sein et non par des décideurs, qu'ils soient individuels ou collectifs.

D'autre part, les volumes de données suivent un accroissement continu pouvant atteindre plusieurs téraoctets pour une société. Bien entendu, ces informations ne se trouvent pas sur un système unique. Prenons l'exemple d'une entreprise quelconque. Elle possède un système de gestion de ses employés basé sur une application développée en interne, un autre système gérant ses achats basés sur un ERP ¹un système de facturation, etc. Chaque système d'information gère un domaine différent de l'entreprise (un « métier » différent) et possède sa propre structure de donnée ainsi que sa propre politique d'accès aux données. Ces données représentent une véritable mine de connaissance sur les produits, les clients, les fournisseurs et les opérations de l'entreprise, etc.

Ceci pose la problématique suivante : Comment prendre des décisions sur la base d'informations issues de systèmes hétérogènes n'ayant pas de moyens pour communiquer facilement entre eux ?

Pour répondre à ces besoins, le nouveau rôle de l'informatique est de définir et d'intégrer une architecture qui serve de fondations aux applications décisionnelles : Un nouveau secteur informatique voit le jour « l'informatique décisionnelle ».

Les outils d'aide à la décision font référence à un ensemble varié d'outils informatiques supportant directement ou indirectement la décision voire le travail général du gestionnaire. Nous nous y intéressons pour trois raisons :

- Ces systèmes vont modifier la décision.
- Afin de produire des systèmes réellement utilisés, il est important d'avoir des modèles justes du décideur et du fonctionnement de la décision.
- On peut considérer qu'un modèle est un système d'aide au travail du chercheur.

2. Décision

2.1. La notion de la décision

La plupart des écoles rationalistes analytiques définissent la décision comme étant un choix entre plusieurs alternatives. Castles et al. (1971), par exemple, attribuent la définition suivante: « *A decision is a conscious choice between at least two possible courses of action* ».

Mintzberg (1973) a défini la décision, qu'elle soit individuelle ou collective, comme « l'engagement dans une action, c'est-à-dire une intention explicite d'agir ». Elle a pour but la résolution de problèmes qui se posent à l'organisation ou à l'individu; et elle peut correspondre à un changement de l'environnement (comportement réactif) ou au désir de saisir une opportunité et ainsi changer l'environnement (comportement d'anticipation). Dans le même sens, Fernandez (1999) a défini la décision par : « *Décider, c'est prendre des risques certes mais c'est être en réaction face à un choix stratégique à prendre* ».

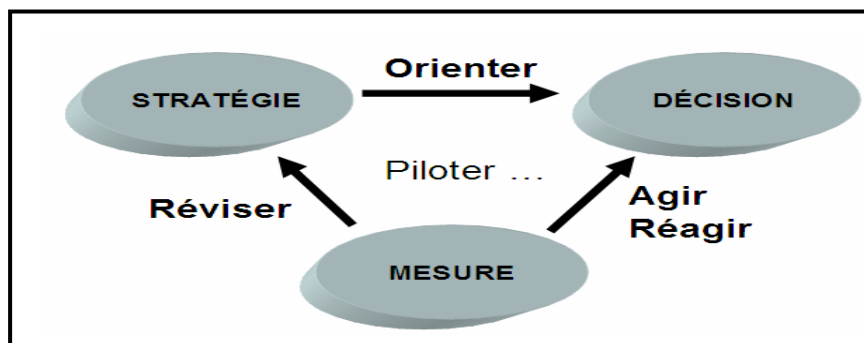


Figure I.1. Lien entre les mesures, la stratégie et la décision. Fernandez (1999)

La figure I.1 parle d'elle-même. Selon Alain Fernandez, pour prendre une décision, il faut pouvoir mesurer. La mesure permettra de réviser la stratégie pour orienter la décision et ainsi d'agir ou de réagir.

Les approches cognitives, de leur côté, traitent la décision comme le résultat d'un processus global de résolution de problèmes (Schneider, 1994). Elles introduisent la notion de temps et de changement en plus des choix. Scott-Morton (1971), par exemple, a distingué pour un processus linéaire de prise de décision entre la découverte de but et la formulation d'objectifs exacts, suivaient de l'étape de sélection de stratégies ou d'alternatives à appliquer permettant d'accomplir ces objectifs; finalement et une fois que les résultats seront atteints, on procède à les évaluer.

La recherche empirique, quand à elle, a remis en cause cette linéarité du processus décisionnel ne reflétant que partiellement la réalité. En effet, des possibilités de faire des retours arrières d'une étape du processus vers une autre située quelque part en avant dans le processus, s'avèrent nécessaires pour la résolution de problèmes réels. Par exemple, Pressman et Weldavski (1984) ou

Bardach (1977) ont tous noté que les buts d'une décision ne seront souvent découverts que durant le processus même de la mise en œuvre de la décision.

2.2. Décideur

La modélisation qui désire savoir comment un décideur, qu'il soit individuel ou collectif, prend une décision doit s'intéresser à la façon dont le décideur modélise le monde et au savoir-faire subjectif et intersubjectif qui permet de traiter cette information.

D'une manière systémique et cognitive, une décision est un processus complexe dont les « données » sont instables et dépendent également de l'interaction du décideur avec son environnement. La notion de décideur isolé est inutilisable dans la plupart des cas. Ainsi, un modèle de décideur devrait contenir également un modèle de l'environnement, même si ce dernier n'est pas sophistiqué. Autrement dit, la modélisation qui désire savoir comment un décideur, qu'il soit individuel ou collectif, prend une décision doit s'intéresser à la façon dont le décideur modélise le monde et au savoir-faire subjectif et intersubjectif qui permet de traiter cette information.

Dans le futur, on espère avoir des modélisations multi facettes du décideur multi incluant, simultanément, les éléments suivants :

- La perception d'une situation en fonction des croyances et d'autres connaissances.
- Les activités de résolution de problèmes y compris des problèmes de choix.
- L'interaction entre décideurs.

2.3. Types de décisions

Revenons sur les étapes de la décision (qui ont donc une signification plutôt logique que temporelle). Celles qui sont les plus souvent citées par les auteurs non réductionnistes sont :

- la perception
- la préparation
- la décision
- l'exécution

Leurs opinions sur l'importance de ces processus varient énormément. Cette variété de vues et de paradigmes s'explique en partie par le type de décision auquel ils s'intéressent. Les éléments ontologiques à l'aide desquels on pourrait définir des types de décision illustrent la richesse de ce concept. Ce sont (Clarke, 1993) :

- Objet de la décision : but, programme, opération, instrument,...
- Organe de décision : organisation, groupe, individu,....
- Type de prise de décision : routine, créatif, application d'un programme,...
- Portée de la décision : stratégique, tactique, opérationnelle,....
- Contrôle des éléments de la décision : bonne, moyenne,...

Sans discuter cette typologie en détail, nous affirmons que la différence structurelle entre certains types de décision peut être très grande. Cela explique en partie la divergence des paradigmes en ce qui concerne la définition de ce qu'est une décision et de ce qu'un décideur sait faire.

Donc, on peut distinguer les types suivants de décisions :

- **Décisions non programmables** : La multiplicité des critères à prendre en compte décourage son utilisation,...
- Exemple* : choix du lieu d'implantation d'une machine à café.
- **Décisions programmables** : Comportent en général peu de paramètre, qui de plus sont le plus souvent aisément identifiables. On peut leur appliquer des procédures formalisées de résolution.
- Exemple* : gestion de stocks, traitement de commandes,...

- **Décisions structurées** : Sont des décisions pour lesquelles les informations nécessaires sont disponibles, les alternatives possibles énumérables et les mécanismes qui permettent de les évaluer connus.
- **Décisions non structurées** : Présentent des caractéristiques spécifiques. On peut distinguer trois types de complexité qui caractérisent ce type de décisions :
 - La complexité géographique (actuellement résolue par l'évolution du monde de la télécommunication et des réseaux).
 - La complexité imprévisible (due à l'incertitude).
 - La complexité dynamique, c'est-à-dire la prise en compte du caractère historique de la décision et de l'émergence. Cette complexité exige que la décision se prenne de façon dynamique.

A l'optique de cette typologie, on peut indiquer que les décisions dans les entreprises se répartissent sur toute la longueur d'une échelle qui va du moins normalisées au plus normalisées.

Les décisions **bien normalisées** reviennent souvent, et c'est probablement pour cette raison que l'organisation a fait l'effort de les normaliser. De nombreuses décisions sont du type normalisé : facturation, achats, détermination des niveaux de production, etc. Quand elle est informatisée, une décision normalisée obéit à un programme dont le déroulement est fixe, il n'y a pas de retour arrière et il n'y a pas d'exploitation heuristique.

Une décision **mal** ou **non normalisée** est prise par une procédure non programmée ou peu programmée, qui reposera probablement sur une stratégie progressive qui comportera des retours arrière. Si on veut formaliser la résolution d'un problème lié à une décision non normalisée, il faudra d'abord construire un espace de résolution, c'est-à-dire structurer et modéliser le problème ; d'où une analogie soulignée par Simon (1977), entre décision mal normalisée et décision mal structurée, et un recouvrement entre les décisions bien structurées et celles qui sont bien normalisées.

Le problème, où la décision paraît bien structurée lors de la formulation du problème ou des actions possibles, s'effectue à l'aide d'un modèle et d'une représentation canonique. La structuration apparaît comme un préalable à une normalisation. Au départ, la décision observée est baptisée spontanée, automatique, intuitive, il s'agit de décisions non structurées. Ensuite, quand on le peut, on passe à la phase de structuration et de modélisation.

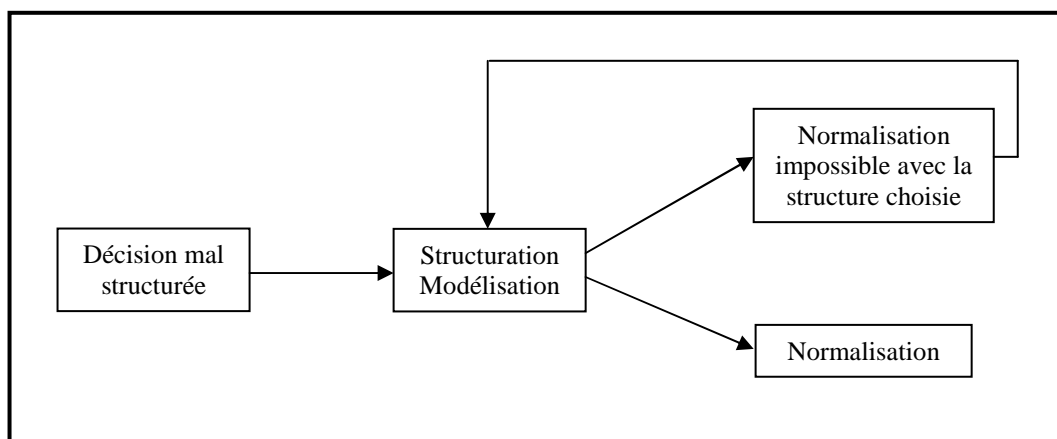


Figure I.2. La structuration intervient avant la normalisation.

Notons que la facilité de normalisation dépend du niveau de la décision : gestion opérationnelle de base, pilotage et planification stratégique.

3. Aide à la décision

L'aide à la décision a été définie par Roy (1993) comme étant : « l'activité de celui qui, prenant appui sur des modèles clairement explicités mais non nécessairement complètement formalisés, aide à obtenir des éléments de réponse aux questions que se pose un acteur dans un

processus de décision, éléments concourant à éclairer la décision et normalement à recommander, ou simplement à favoriser, un comportement de nature à accroître la cohérence entre l'évolution du processus d'une part, les objectifs et le système de valeurs au service desquels ces acteurs se trouvent placés d'autre part ».

4. Le décisionnel

Le décisionnel est défini par : « *Le décisionnel est le processus visant à transformer les données en informations et, par l'intermédiaire d'interrogations successives, transformer ces informations en connaissance* ».

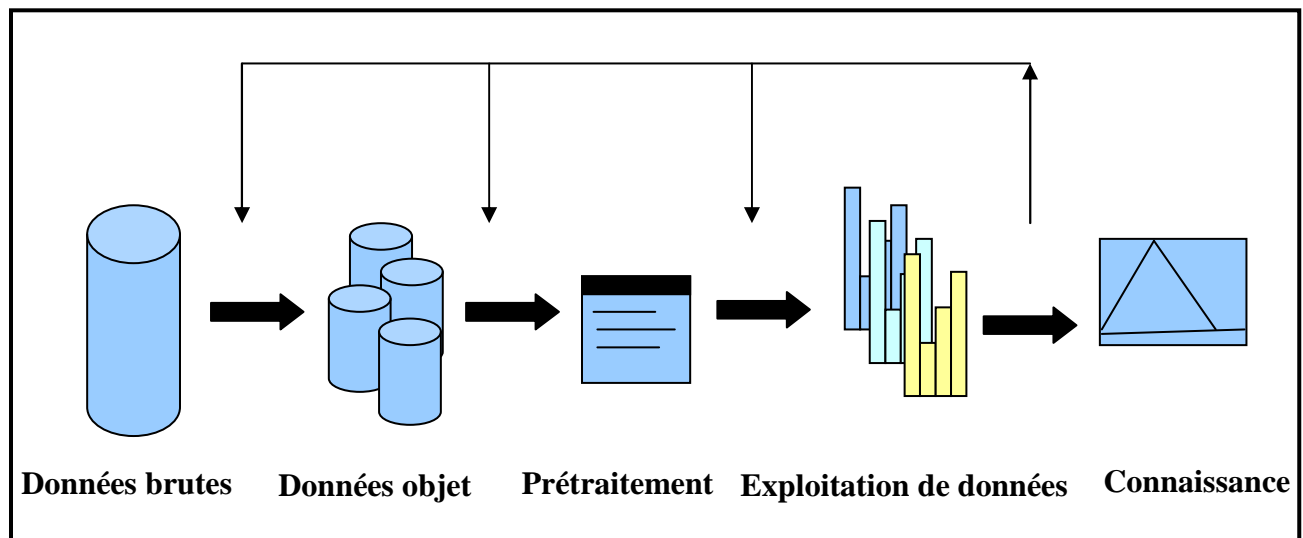


Figure I.3. Procédé d'extraction des connaissances.

5. Le processus de décision

La décision ne se réduit pas à un simple choix à faire entre plusieurs alternatives. SIMON décrit un processus itératif dans le modèle I.D.C (Intelligence Design Choice) qui découpe le processus de décision en trois phases pour :

- Identifier dans l'environnement les facteurs que l'on considère sensibles.
- Organiser et modéliser les informations que le décideur a retenues de façon à disposer des solutions possibles pour résoudre son problème.
- Permettre au décideur de faire un choix parmi les solutions établies ou encore de faire une itération sur l'une des phases précédentes.

En effet, s'il n'a pas trouvé de solution satisfaisante après ces trois phases, on peut ajouter une phase de contrôle qui peut être mise en œuvre avant (à travers des solutions) ou après l'application de la décision.

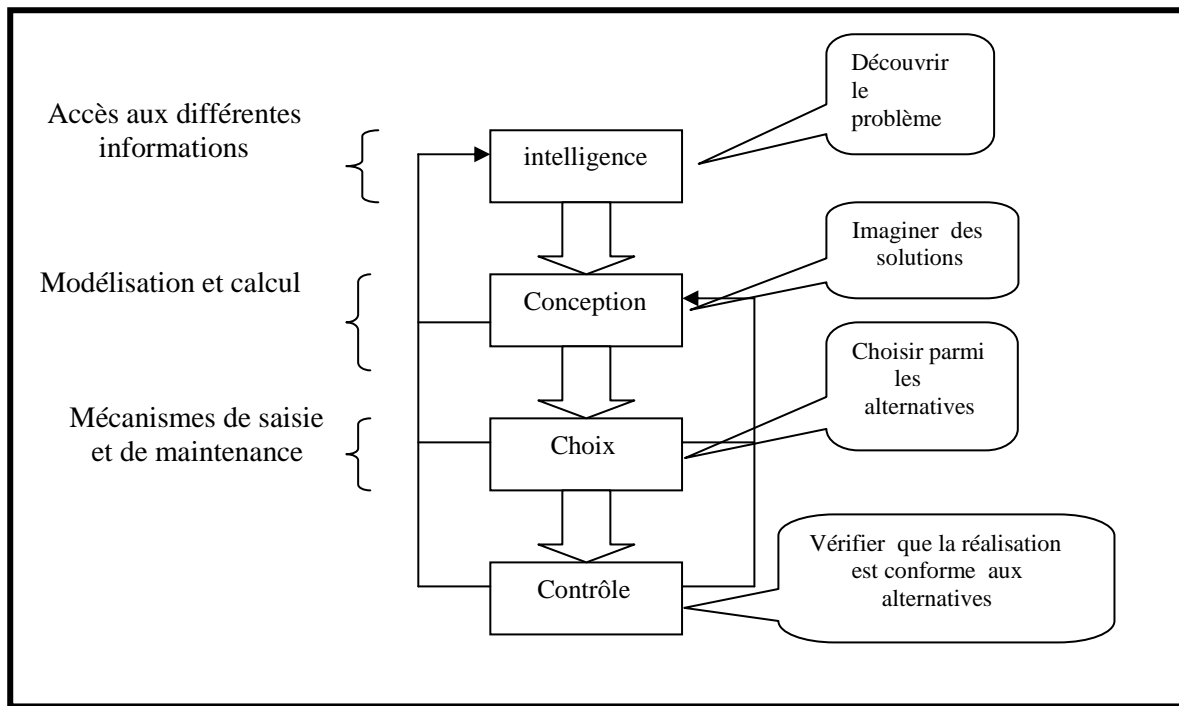


Figure I.4. Le modèle I.D.C. (Marakas, 1999)

Simon (Simon, 1977) distingue trois phases dans le processus de décision :

- **L'information ou le renseignement (intelligence)** : Recherche d'information en fonction des questions que se pose le décideur, définition du problème à résoudre, c'est-à-dire identification des objectifs ou buts, classification de celui-ci, décomposition du problème en sous-problèmes – beaucoup de problèmes complexes peuvent être divisés en sous-problèmes plus simples à résoudre qui aide à résoudre le problème plus complexe.
- **La conception (design)** : la construction ou le choix d'un modèle du processus de décision, détermination des variables de décision, des variables incontrôlables, des relations mathématiques ou symboliques ou qualitatives entre ces variables et construction de solutions.
- **Le choix (choice)** : Le décideur choisit entre les différentes suites d'actions-solutions- qu'il a été capable de construire et d'identifier pendant la phase précédente. Cette étape se décompose en deux étapes : celle de **recherche** et celle d'**évaluation**. La phase de recherche peut être de type *analytique* (optimisation, toutes les alternatives sont atteintes), *aveugle* (recherche exhaustive ou partielle) et heuristique. Dans ces deux cas, seule une solution respectivement soit assez bonne, soit satisfaisante est fournie. Cette recherche est couplée avec une évaluation qui est l'étape finale qui amène à une recommandation solution. Parmi les méthodes utilisées, on peut citer : méthodes pour buts multiples, analyse de sensibilité, méthode « what-if », méthode « goal seeking ».

Ce processus n'est pas obligatoirement séquentiel, il peut y avoir des retours arrière. L'acquisition d'informations pertinentes peut se poser en terme de décision. Elles sont en effet à l'origine du processus et donc influencent fortement les phases suivantes. Il est devenu classique de distinguer les « programmed decisions » et les « non-programmed decisions », traduit ici par « décision bien structurée » et « décision peu structurée ». En fait il existe un continuum allant des moins structurées aux plus structurées. Une décision est bien structurée quand un processus connu et explicite existe permettant de traiter les informations dans le système (Lévine et Pomerol 1989). Elle correspond à un programme immuable et fixe.

De nombreux problèmes dans les organisations peuvent s'analyser en terme d'allocation de ressources (argent, temps, pouvoir, personnes, espace, équipement, l'affectation d'employés ou d'équipements à des travaux,...). Un décideur doit allouer des ressources peu abondantes pour des activités variées afin d'optimiser un objectif mesurable. L'objectif est dans ce cas de minimiser ou

maximiser une variable (l'objectif mesurable) : minimisation d'un coût ou maximisation de l'occupation de l'espace ou du temps. Les techniques d'optimisation via la programmation mathématique (**programmation linéaire** par exemple) sont ici utilisées, c'est-à-dire qu'il y a plusieurs variables reliées par une équation mathématique et soumises à des contraintes, il s'agit de maximiser l'une d'entre elles. La solution est la meilleure solution possible. Ces techniques supposent que : les résultats des différentes allocations peuvent être comparées et indépendantes, le résultat total –variable à maximiser ou minimiser- est égale à la somme des résultats des différentes activités, toutes les données sont connues et certaines, les ressources sont à utiliser de la manière la plus économique. Lorsqu'il y a plusieurs variables (objectifs)- souvent contradictoires- à optimiser, l'**analyse multicritères** est utilisée.

6. Historique des systèmes décisionnels

Les primitives de l'informatique décisionnelle se sont développées dans les années 70. Historiquement, les grandes firmes ont été les premières à comprendre la valeur ajoutée des outils d'aide à la décision. En effet, disposant de quantités considérables d'informations dans leurs bases de données opérationnelles, elles ont, en premier lieu, commencé par les interroger directement, par des requêtes. Cette solution de remplacement a vite montré ses limites aussi bien en temps qu'en ressources humaines et matérielles. De nouveaux outils apparaissent alors. Ces outils ont connu l'évolution suivante :

- **L'infocentre** : L'infocentre est une copie à l'identique des bases de données dans un nouvel environnement (années 70 et 80).
- **SBC** : Systèmes à Base de Connaissances
- **L'EIS** : Executive Information System proposant les premiers *tableaux de bord* dans les années 1990.
- **Les entrepôts de données** : Les entrepôts de données (DW) sont considérés comme étant le lieu de stockage des gros volumes de données.
- **Les bases de données multidimensionnelles** : Une base où chaque indicateur est analysé en fonction de plusieurs critères ou dimensions.
- **La Business Intelligence** : Qui regroupe les fonctions d'analyse des données et de reporting.

Outre l'évolution des architectures, l'analyse des données dans ces systèmes d'information décisionnels a également connu une évolution.

L'*informatique décisionnelle* est alors essentiellement constituée d'outils d'édition de rapports, de statistiques, de simulation et d'optimisation. Provenant des recherches en *Intelligence Artificielle*, les *systèmes experts* voient le jour. Ils sont conçus par extraction de la connaissance d'un ou plusieurs experts et sont des systèmes à base de règles. De bons résultats sont obtenus pour certains domaines d'application tels que la médecine, la géologie, la finance, etc. Cependant, il apparaît vite que la formalisation sous forme des règles de la prise de décision est une tâche difficile voir impossible dans de nombreux domaines.

Dans les années 90, deux phénomènes se produisent simultanément :

- La possibilité de conception d'environnements spécialisés pour l'aide à la décision.
- De nombreux algorithmes permettant d'extraire des informations à partir de données brutes arrivent à maturité. Ces algorithmes ont des origines diverses et souvent multiples. Certains sont issus des statistiques ; d'autres proviennent des recherches en Intelligence Artificielle, recherches qui se sont concentrées sur des projets moins ambitieux, plus ciblés; certains s'inspirent de phénomènes biologiques ou de la théorie de l'évolution. Tous ces algorithmes sont regroupés dans des logiciels de *fouille de données* ou *Data Mining* qui permettent la recherche d'informations nouvelles ou cachées à partir de données. Ainsi, dans le cas de systèmes à base de règles, plutôt que d'essayer d'extraire la connaissance d'experts et d'exprimer cette connaissance sous forme de règles, un logiciel génère ces règles à partir de données.

7. Références bibliographiques

- Bardach E., The implementation game : what happens after a bill becomes a law. The MIT Press, Cambridge, MA, 1977.
- Castles F.G., Murray D.J. et Potter D.C., Decision, Organisations and Society, Penguin, Harmondsworth, U.K., 1971.
- Clarke A. et Smith M. G., A cooperative computer based on the principles of humans cooperation. International Journal of Man-Machine Studies (IJMMS), 1993.
- Fernandez A., Les nouveaux tableaux de bord des managers. Groupe Eyrolles, 1999.
- Le Moigne, J.L., Les systèmes de décision dans les organisations. Presses Universitaires de France, 244 p, 1974.
- Kahnemann D., Slovic P. et al. Judgment under Uncertainty : Heuristics and Biases, Cambridge University Press, Kahnemann D., Slovic P. et al., Eds., 1982.
- Lévine P. et Pomerol J.C, Systèmes interactifs d'aide à la décision et systèmes experts, Edition Hermès, Paris, 1989.
- Marakas G.M., Decision Support Systems in the Twenty-first Century. Prentice-Hall : Upper Saddle River, NJ, xxi, 506, 1999.
- Mintzberg H., Strategy-making in three modes. California Management Review 16 (2) : 44-53, 1973.
- Pressman J. L. et Wildavsky A., Implementation : how great expectations in Washington are dashed in Oakland : or, why it's amazing that federal programs work at all : this being a saga of the economic development administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes. Univ. of California Press, Berkeley, 1984.
- Roy, B. et D. Bouyssou D., Aide multicritère à la décision : méthodes et cas. Economica, Paris, 1993.
- Schneider D.K., Modélisation de la démarche du décideur politique dans la perspective de l'intelligence artificielle. Thèse de doctorat, université de Genève, 1994.
- Scott-Morton M., Management decision systems, computer based support for decision making, Harvard University, Boston, 1971.
- Simon H., The New science of management decision, Prentice hall, Englewood-Cliffs, 1977.
- Voir aussi :**
- Adam. F. Comment les Managers et les organisations prennent des décisions et implications pour l'Aide à la Décision. Université Pierre et Marie Curie (UPMC), 2008.
- Adam. F. and Pomerol. J-Ch. How Decision Making and Knowledge representation evolved in an organization after computerisation. Decision Making Support Systems: Achievements and Challenges for the New Decade, 2008.
- Hadoux. E., Beynier. A. and Weng. P. Prise de décision séquentielle en environnements incertains et non stationnaires. 15ème Congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF 2014), Bordeaux, 2014.
- Lauritzen. S. and Nilsson. D. Representing and Solving Decision Problems with Limited Information. Management Science, 47(9):1235–1251, 2001.
- Mintzberg, H., Westley, F., Decision making : it's not what you think. MIT Sloan Management Review 42 (3) : 89-93, 2001.
- Pomerol. J-Ch. Decision-Making and Action. ISTE, Wiley, 2012.

Shim J.P., Warkentin M., Courtney J., Power D., Sharda R., Carlsson C. Past, present, and future of decision support technology. *Decision Support Systems*, Elsevier, volume 33, pp 111-126, 2002.

Chapitre II : Systèmes à base de connaissances

1. Introduction

Un jour, quelque part, nous arrivons à un carrefour où des routes se croisent et où ne figure aucune indication. Plusieurs solutions pour trouver le bon chemin s'offrent à nous : choisir une route au hasard, par exemple continuer tout droit coûte que coûte, ou plutôt choisir une route en s'appuyant sur des connaissances disponibles sur le domaine, par exemple avec l'aide d'une carte de la région, d'un guide humain ou matériel, en repérant notre position par rapport au soleil ou avec une boussole, etc. Beaucoup de problèmes se présentent comme celui-ci, où il s'agit de partir d'un état initial pour atteindre un état final, en passant par un ensemble d'états intermédiaires qu'il faut explorer et à partir desquels il faut faire des choix en fonction de connaissance plus ou moins disponibles et plus ou moins complètes. Le nombre de possibilités est la plupart du temps très vaste, si bien qu'il n'est pas envisageable de choisir au hasard, sous peine de ne jamais résoudre le problème. La problématique des systèmes à base de connaissances repose sur de tels constats : comment procéder pour qu'un système informatique s'appuyant sur des connaissances puisse résoudre des problèmes dans un domaine donné ? Cela revient, pour un tel système, à faire preuve d'une certaine forme d'intelligence, à être capable de mettre en œuvre des stratégies et à faire des choix judicieux lorsque c'est nécessaire. Les connaissances dont il est question pour ces systèmes sont disponibles et exploitées sous forme électronique. Ainsi, la tâche principale d'un système à base de connaissances d'un domaine pour aider à résoudre un problème donné, d'identification, de classification, de reconnaissance, de diagnostic, de configuration, de planification, etc. pour en revenir à notre voyage initial, de nombreux problèmes peuvent se poser en cours de route qu'il peut s'avérer fort utile de résoudre, comme identifier un animal qui vient à passer, dangereux ou pas, soupçonner ou détecter l'existence d'un point d'eau, diagnostiquer une panne de la voiture et réparer, trouver un chemin plus court ou plus pratique, ou encore plus long mais plus joli, ...

2. La notion d'un système à base de connaissances

2.1. La résolution de problèmes

Pour résoudre un problème, un être humain raisonne généralement sur des concepts abstraits qui modélisent les objets de l'univers du problème, en tire des conclusions qu'il interprète ensuite dans cet univers. La simulation d'un tel comportement pour un système informatique se décompose selon les étapes suivantes :

- L'abstraction : permet d'associer des structures aux éléments du domaine et du problème considéré ;
- Les structures sont ensuite représentées en machine sous forme d'expressions symboliques (formules bien formées) à l'aide d'un langage de représentation des connaissances. Parmi ces expressions, certaines sont permanentes et constituent la mémoire à long terme ; d'autres sont temporaires et constituent la mémoire à court terme. La mémoire à long terme contient la connaissance dont dispose le système sur le domaine étudié et peut être considérée comme un modèle opérationnel et utilisable du domaine, ce qui s'appelle encore une

ontologie. La mémoire à court terme contient les informations sur le problème particulier à résoudre ;

- Les expressions symboliques sont combinées entre elles pour produire, par application de règles d'inférence, de nouvelles expressions symboliques ;
- Les nouvelles expressions symboliques sont interprétées dans le cadre de l'univers du problème considéré et apportent des éléments de solution au problème courant.

2.2. Connaître c'est décrire pour retrouver

Concevoir un système à base de connaissances consiste à modéliser puis à représenter des connaissances propres à un domaine à l'aide d'un formalisme de représentation, puis à manipuler les connaissances par l'intermédiaire de règles d'inférence pour résoudre les problèmes posés. La connaissance est alors considérée comme une *entité calculable*, ce qui pose un certain nombre de problèmes pratiques voire philosophiques, en particulier :

- Sous quelle forme doivent s'exprimer les connaissances ?
- Comment un mécanisme de raisonnement peut-il exploiter au mieux un ensemble de connaissances forcément limité et comment peut-il tirer de cet ensemble toutes les connaissances implicites qu'il renferme ?
- Comment les connaissances inférées influent-elles sur le comportement du système ?
- Comment raisonner avec des informations incomplètes ou bruitées ?
- Comment faire aboutir un raisonnement alors que l'éventail des possibilités de recherche d'une solution est virtuellement infini ?

Une représentation des connaissances se caractérise par une *syntaxe* et une *sémantique* (Guenther, 1987) : les expressions du langage dérivent d'un procédé de construction de formules bien formées (syntaxe) et se voient associées une sémantique, sur laquelle repose la validité des inférences effectuées par le système. La sémantique est généralement donnée par une fonction d'interprétation qui met en correspondance les éléments de la représentation et les éléments d'un domaine d'interprétation (la sémantique dénotationnelle). Les opérations de manipulation de connaissances essentiellement recherche et inférence sont dépendantes du formalisme de représentation lui-même. Les concepteurs d'une base de connaissances doivent choisir les unités de connaissances du domaine qu'ils représentent et les coder de façon à ce que la base construite soit un modèle opérationnel de l'univers considéré.

La plupart des formalismes de représentation des connaissances sont *déclaratifs* : les connaissances expriment ce que l'on sait, le « quoi » de l'univers étudié, par opposition à un formalisme *procédural*, où est explicité le « comment » des choses. L'aspect déclaratif repose sur une certaine indépendance vis-à-vis du contexte d'utilisation – il est possible d'appréhender les unités de connaissances isolément et chaque unité est intelligible pour et par elle-même – et vis-à-vis des traitements ; il n'est pas nécessaire de « faire tourner le système » pour savoir ce qu'il renferme et ce qu'il censé faire.

3. L'architecture et la mécanique des systèmes de connaissances

3.1. Les concepts et les ontologies

En représentation de connaissances, le terme *ontologie* fait référence à un modèle opérationnel utilisé pour décrire un domaine particulier du monde réel. Dans cet ordre d'idées, les *concepts* apparaissent comme des briques de base des ontologies : ils possèdent une *intention* qui se définit par l'ensemble des propriétés caractéristiques, ou attributs, du concept et une *extension* qui recouvre l'ensemble des individus ou objets « instances » du concept. L'intension peut s'appréhender comme l'ensemble des conditions nécessaires et suffisantes devant être vérifiées par un objet pour être instance du concept. Les intentions et les extensions sont emboîtées en sens inverse l'une de l'autre : plus un concept est général plus il recouvre d'individus et moins il recouvre d'individus et moins il recouvre de propriétés, et réciproquement.

Pratiquement, une ontologie se présente comme un système formel constitué d'un ensemble de concepts et d'un ensemble de relations binaires spécifiées par des couples de concepts de *domaines* et de *codomaines*, d'un ensemble d'axiomes, d'une relation de *spécialisation* qui est généralement réflexive, antisymétrique et transitive, qui permet d'organiser les concepts et les relations en une hiérarchie, et qui autorise les inférences. Il est possible d'inférer qu'un individu est instance d'un certain concept ou qu'un concept partage certaines propriétés avec un autre concept. Suivant ce schéma, une base de connaissances s'appuie sur un couple constitué d'une ontologie et d'une base d'assertion ou de faits (dans lesquels interviennent les individus). Ainsi, une base de connaissances contient des unités de formes et de niveaux d'abstraction différents, comme par exemple des concepts, des instances, mais aussi des règles manipulant des faits, des stratégies (*heuristiques*) exprimant la façon de se servir des connaissances de base.

Les ontologies (Staab, 2004), (Uschold, 1996) ont une place d'importance croissante dans des domaines comme la gestion des connaissances (Dieng-Kuntz, 2004), l'intégration et la recherche d'information, le commerce électronique (Ontoweb, 2002) et bien sûr le Web sémantique (Bechhofer, 2001), (Sure, 2002). Les ontologies sont appelées à jouer là un rôle clé en établissant une terminologie commune entre les agents – logiciels et humains – qui peuvent ainsi partager la même sémantique sur les concepts et les relations manipulés. Les ontologies sont également au cœur de la gestion des connaissances, où, à l'image de la gestion de base de données, ce ne sont plus simplement des données (syntaxiques) mais des connaissances (munies d'une syntaxe et d'une sémantique) qui sont considérées et manipulées.

Toutefois, les objets du quotidien obéissent rarement à des lois rigoureuses, comme les êtres humains et au contraire des objets mathématiques. Intégrer les différentes natures des objets du quotidien dans un formalisme de représentation pose des problèmes difficiles à résoudre comme la représentation de modalités (statut et degré de vérité des informations), la représentation de connaissances typiques et exceptionnelles, la représentation de connaissances incomplètes, évolutives, interdépendantes, etc. Ainsi, de nombreux formalismes de représentation et de raisonnement ont été mis au point pour prendre en compte les natures diverses et variées des connaissances. Quelques-uns de ces formalismes parmi les plus couramment utilisés, les logiques des propositions et des prédicats, le langage Prolog, les règles de production et les logiques de descriptions, sont présentés dans les paragraphes qui suivent. Il existe bien sûr d'autres formalismes de représentation, comme par exemple les graphes conceptuels ou les représentations de connaissances par objets.

3.2. L'ingénierie et la gestion des connaissances

L'ingénierie d'un système à base de connaissances s'appuie sur un ensemble d'opérations qui se retrouvent dans la conception de tout logiciel d'envergure :

- **Initialisation et modélisation** : ces étapes recouvrent la spécification d'un modèle des éléments du domaine à représenter et une mise en œuvre du modèle, des méthodologies de modélisation telles que KADS (Martin, 1994) et CommonKADS ont été mises au point pour ces besoins ;
- **Représentation et raffinement** : ces étapes recouvrent la phase de représentation proprement dite, le choix d'un langage de représentation, l'implantation du modèle et le raffinement du modèle après les premières opérations de test ;
- **Evaluation** : cette étape suit et complète l'utilisation du système à base de connaissances dans les applications et recouvre la mise en place d'environnements logiciels adaptés aux besoins spécifiques et l'évaluation du fonctionnement du système dans la pratique ;
- **Maintenance** : les connaissances évoluent, les modes de raisonnement aussi, ce qui fait évoluer d'autant la spécification du système, qui doit être mise à jour pour garantir la cohérence et la compatibilité des connaissances, anciennes et nouvelles ;
- **Diffusion** : les connaissances doivent être partagées et transmises si besoin est sous une forme opérationnelle : c'est là un des fondamentaux de la conception des ontologies et donc

des bases de connaissances. Le fait que les connaissances soient codées à l'aide d'un langage de représentation muni d'une syntaxe et d'une sémantique bien définies garantit que ce sont bien les mêmes éléments de connaissances qui sont envoyés et réceptionnés, au sens où ils peuvent servir à résoudre des problèmes de même nature de part et d'autre.

Du côté du monde industriel, ce sont les notions de gestion des connaissances et de mémoire d'entreprise qui émergent effectivement. Les problèmes à résoudre consistent pour l'essentiel à gérer les connaissances liées à l'entreprise : les recenser, les mémoriser, les utiliser, les transmettre et les faire croître. Les éléments essentiels de la mémoire d'entreprise sont le savoir-faire, l'expertise, les documents scientifiques et techniques ; les différents « agents » sont ici les personnes, les connaissances, les documents et les actions (la dynamique, le flot des informations, etc). Un des problèmes récurrents qui se pose est celui de « trouver la bonne information » : qui sait ou peut savoir et sinon comment faire.

4. Les formalismes logiques et Prolog

4.1. Le calcul des propositions

Noté CP0, correspond à la logique où les constituants des formules sont des propositions, c'est-à-dire des expressions qui prennent l'une des deux valeurs {faux, vrai} (ou {0,1}), à l'image d'une fonction booléenne. Une proposition peut correspondre à une forme du langage comme *sujet-verbe-complément*, insécable au sens où elle forme un « tout non décomposable » : il-fait-beau ou le-chat-est-noir. Le calcul des propositions permet de formaliser des énoncés de problèmes donnés en langage naturel et une certaine partie du raisonnement humain. Les problèmes principaux qui se posent en CP0 sont :

- P1)** de représenter des énoncés avec des formules du CP0,
- P2)** de savoir si une formule est toujours vraie,
- P3)** et de savoir si une formule se déduit d'un ensemble de formules.

Ces problèmes recouvrent les tâches qui doivent être accomplies par un système à base de connaissances.

4.2. Le calcul des prédicats du premier ordre

Abrégé en CP1, étend le CP0 avec des variables (à valeurs non nécessairement booléennes), des fonctions qui retournent des valeurs (non booléenne), des prédicats qui généralisent les propositions et des quantificateurs qui déterminent la portée des variables. Par exemple, le prédicat unaire *masculin(x)* prend la valeur *vrai* chaque fois que la variable *x* désigne un individu de sexe masculin ; le prédicat binaire *est-mère-de(x,y)* prend la valeur *vrai* chaque fois que la variable *x* désigne la mère de l'individu désigné par la variable *y*. Le calcul des prédicats devient indispensable dès lors qu'il est nécessaire de représenter des formules où interviennent des quantifications, pour décrire par exemple des propriétés qui concernent un ou tous les individus étudiés.

En particulier, la représentation de phrases du langage naturel s'appuie sur des formes génériques universelles et existentielles, où interviennent explicitement les quantificateurs. Ainsi, dans le célèbre syllogisme *tous les hommes sont mortels, Socrate est un homme, donc Socrate est mortel*, la première phrase illustre une quantification universelle, la deuxième une instanciation et la troisième une conséquence logique des deux précédentes. C'est à représenter ce type d'énoncés avec des formules et à prouver qu'une formule est une conséquence logique d'un ensemble de formules que s'attache le calcul des prédicats. Les deux premières phrases du syllogisme se traduisent par les deux formules : $(\forall x)(\text{homme}(x) \rightarrow \text{mortel}(x))$, $\text{homme}(\text{Socrate})$; En appliquant la règle de résolution – comme pour le CP0 – moyennant la substitution de la variable *x* par *Socrate*, il est possible de déduire que $\text{mortel}(\text{Socrate})$ est une conséquence logique des deux premières phrases.

4.3. Prolog

Par rapport aux systèmes à base de connaissances, le CP1 permet l'utilisation de variables et donc de traiter une catégorie de problèmes plus vaste que ne le permet le CP0. En contrepartie, la modélisation des énoncés et le raisonnement s'avèrent plus complexes. Du point de vue pratique, l'utilisation du CP1 pour concevoir un système à base de connaissances repose le plus souvent sur le langage de programmation logique **Prolog** (Roussel, 1975), (Colmerauer, 1996), , qui adopte les principes mêmes du CP1 et qui a été mis au point dans les années soixante-dix. Les formules du langage Prolog se composent de constantes, de variables, de fonctions, de prédicats – où certaines fonctions et certains prédicats sont prédéfinis – et où sont également disponibles des structures de données comme la liste par exemple.

Un programme Prolog se présente comme une suite de faits et de règles, qui correspondent à la base de faits et à la base de connaissances. L'utilisation d'un système Prolog consiste à poser des questions et à obtenir des réponses par l'intermédiaire d'une instanciation des variables qui figurent dans les questions. Pour satisfaire la question $?q(x)$, le système Prolog recherche une formule (fait ou règle) de tête $q(.)$ avec laquelle le prédicat de la question peut s'apparier, puis il retourne la substitution des variables qui autorise l'appariement. Par exemple, $?plat(p)$ peut s'apparier avec $plat(couscous)$, ce qui produit la réponse $\{p = couscous\}$. Lorsque la tête d'une règle s'apparie avec une question, la tête s'efface et les prédicats du corps de la clause deviennent à leur tour autant de questions, dans lesquels les substitutions de variables effectuées lors de l'appariement se sont propagées. Le processus se continue jusqu'à ce que tous les faits et toutes les règles aient été traités. Par exemple, la question $?repas(e,p,d)$ donne naissance à trois sous-questions qui sont $?horsDoeuvre(e)$, $?plat(p)$ et $?dessert(d)$, qui donne la réponse $\{e = carotte, p = couscous, d = pomme\}$. Le processus est en échec si un appariement n'est pas possible et qu'une tête de règle ne peut pas s'effacer (le système effectue un retour arrière si nécessaire), sinon le processus retourne la suite des substitutions associées à tous les appariements réalisés.

Le langage Prolog a eu une importance assez considérable dans le développement de systèmes à base de connaissances et reste encore et toujours un langage d'actualité pour l'ingénierie des connaissances. Des expériences ont également été menées pour coupler Prolog avec des langages de contraintes, ce qui a accru le potentiel de du langage en termes de conception de systèmes à base de connaissances.

5. Les systèmes à base de règles

Les systèmes à base de règles (de production), plus connus sous le nom de **systèmes experts** (Jackson, 1998), constituent un moyen privilégié de coder des connaissances d'experts. En effet, il s'avère qu'une bonne partie des éléments d'expertise courants et utiles en résolution de problèmes, par exemple, pour une analyse, un diagnostic, une reconnaissance, une classification, s'exprime sous la forme de règle de production. Fondamentalement, les inférences réalisées par les systèmes à base de règles sont régies par les principes de la logique, mais c'est le recueil d'expertise qui fait ici la différence. Une règle de production se compose partie gauche – la prémisse- et d'une partie droite – la conclusion – et se présente sous la forme *si condition(s) alors conclusion(s)*. Le langage de description des règles peut être plus ou moins complexe, comporter des conjonctions, des disjonctions, des négations, des variables, etc.

Un système à base de règles se compose de trois modules séparés et interdépendants. La base de règles contient les connaissances du domaine mémorisées sous forme de règles de production. Les règles sont intelligibles pour et par elles-mêmes, et elles sont indépendantes les unes des autres (aucune règle de production ne fait appel à une autre règle). La base de faits relatifs au problème courant à résoudre. Le moteur d'inférence permet d'enchaîner et de contrôler les cycles d'applications des règles, en partant des faits pour atteindre la solution du problème courant. Le cycle d'un moteur d'inférence est à trois temps : sélection des règles applicables, choix d'une règle et déclenchement.

6. Histoire, applications et enjeux actuels

L'histoire des systèmes à base de connaissances n'est pas si mais elle a vu naître au bon nombre de systèmes dont certains ont été et sont encore des modèles du genre dans leurs spécialités. Parmi ces systèmes, beaucoup sont restés des prototypes et ont été construits plus pour valider des idées et des résultats de recherche que pour être véritablement opérationnels dans le monde industriel. Parmi les tâches effectuées par ces systèmes, figurent la classification, la configuration et le diagnostic. Les domaines d'application sont très variés pour les trois tâches, avec, aux premiers rangs, la biologie, la chimie, la géologie, la médecine, mais aussi l'agronomie, l'environnement, le marketing, les mathématiques, la sidérurgie, etc.

Pour la classification qui recouvre l'identification d'un objet ou phénomène quelconque en tant que membre d'une classe connue, les systèmes qui ont forcément marqué leur époque sont MYCIN (médecine), DENDRAL (chimie) et PROSPECTOR (géologie).

Dans le contexte actuel, plusieurs défis et enjeux se posent pour les systèmes à base de connaissances, surtout par rapport à la problématique du Web sémantique. La diffusion des connaissances est essentielle et ne peut pas se faire sous n'importe quelle forme : ce sont les ontologies qui vont donner une forme et un sens aux connaissances représentées, qui vont servir de véhicules et qui vont permettre à des agents de tirer parti des connaissances pour résoudre des problèmes, manipuler des connaissances distribuées, rechercher et exploiter des documents en fonction de leur contenu, assurer l'interopérabilité entre différents services, etc. Pour l'heure, il est acquis que la structure et la sémantique des documents – informations, connaissances – doit être diffusée par l'intermédiaire de langage de représentation comme OWL, compatible avec les standards du Web qui sont devenus XML et RDF, pour la description de documents et des ressources. Il est nécessaire pour les systèmes à base de connaissances d'avoir une syntaxe et une sémantique clairement définies, et de montrer des possibilités de représentation et de raisonnement réelles et efficaces. Est-ce qu'un « super système à base de connaissances universel », dans la lignée de CYC (), verra bientôt le jour, avec un ensemble de dimensions spécifiques, une pour décrire les objets du monde, une pour les documents, une pour les structures de données, une pour extraire des connaissances, une pour les traitements et les services, une pour les raisonnements de toutes natures ? Dans cette perspective de « gratte-ciel de connaissances », chacun des étages a son rôle à jouer et doit collaborer avec les autres – via l'ascenseur ou l'escalier – pour résoudre les problèmes posés par les visiteurs ...

7. Références bibliographiques

- Bechhofer S., Horrocks I., Goble C. & Stevens R., OilEd: a Reason-able Ontology Editor for the Semantic Web. In Joint German/Austrian conference on Artificial Intelligence (KI'01), volume (2174) of Lecture Notes in Artificial Intelligence, p. 396–408, Vienne, Autriche: Springer Verlag, 2001.
- Colmerauer A. et Roussel Ph., The birth of Prolog, 1992. Published in Book : History of programming languages - II, pp 331-367, ACM, New York, 1996.
- Dieng-Kuntz R., Minier D., Corby F., Corby O., Alamarguy L. et Luong P-H., Exploitation d'ontologies pour la gestion des connaissances et le travail collaboratif dans un réseau de soin. Projet ACACIA, INRIA, Sophia Antipolis, 2004.
- Guenther F. et Sabatier P., Sémantique formelle et représentation des connaissances. Revue LANGAGES, vol 22, num 87, pp 103-122, 1987.
- Jackson P., Introduction To Expert Systems. Addison-Wesley, 3 édition, 1998.
- Martin, P., La Méthodologie d'acquisition des connaissances KADS et les explications, Rapport de recherche N. 2179, Paris, INRIA, 1994.

Ontoweb C., Ontology-based information exchange for knowledge management and electronic commerce. Deliverable 1.3, IST Project IST- 2000-29243 OntoWeb, 2002.

Roussel Ph., Prolog, manuel de référence et d'utilisation, Groupe Intelligence Artificielle, Faculté des Sciences de Luminy, Université Aix-Marseille II, France, 1975.

Staab S., Studer R. Eds., Handbook on Ontologies. International Handbooks on Information Systems. Springer, 2004.

Sure Y., Erdmann M., Angele J., Staab S., Studer R. & Wenke D., OntoEdit: Collaborative Ontology Engineering for the Semantic Web. In I. HORROCKS & J. HENDLER, Eds., First International Semantic Web Conference (ISWC'02), volume (2342) of Lecture Notes in Computer Science, pp. 221–235, Chia, Sardaigne, Italie: Springer Verlag, 2002.

Uschold M. & Gruninger M., Ontologies: Principles, methods and applications. Knowledge Engineering Review, 1996.

Voir aussi :

Brézillon. P. From expert systems to context-based intelligent assistant systems: a testimony. The Knowledge Engineering Review, 26:1, pp 19-24, 2011.

Soubie J.L. Coopération et Systèmes à base de connaissances. Habilitation à diriger des recherches, Université Paul Sabatier, Toulouse, 1996.

Teulier-Bourguine R. et Zaraté P. Vers une problématique de l'aide à la décision utilisant les connaissances. Actes de la Conférence en Ingénierie des Connaissances IC'01, Grenoble, France, pp 147-166, 2001.

Chapitre III : Les Systèmes décisionnels

1. Le besoin d'un système décisionnel

Le développement des premiers systèmes d'informations s'est concentré sur l'automatisation des processus opérationnels, ainsi que sur les données liées aux processus

1.1 Le système d'information

« Le système d'information est l'ensemble des méthodes et moyens recueillant, contrôlant et distribuant les informations nécessaires à l'exercice de l'activité en tout point de l'organisation. Sa fonction est de produire et de mémoriser les informations, représentation de l'activité du système opérant (système opérationnel), puis de les mettre à disposition du système de décision (système de pilotage) ».

Cependant ces systèmes ne sont pas adaptés pour faire des analyses complexes de données et ne préparent pas les données pour la prise de décision. Pour assurer une plus grande réactivité et une plus grande compétitivité, les approches traditionnelles s'avèrent donc rapidement insuffisantes et les décideurs requièrent des systèmes basés sur la technologie OLAP qui facilite leur processus de prise de décision d'où l'émergence d'une nouvelle catégorie du système dit : **système décisionnel**.

1.2. Système d'information décisionnel

« Un système d'information décisionnel, SID, est un système qui réalise la collecte, la transformation des données brutes issues de sources de données et le stockage dans d'autres espaces ainsi que la caractérisation des données résumées en vue de faciliter le processus de prise de décision ».

Le système d'information décisionnel doit remplir trois fonctions :

- L'extraction de données.
- Le stockage.
- La restitution des données sous une forme exploitable.

2. Outils d'aide à la décision

2.1. Typologie des outils d'aide à la décision

2.1.1 Classement en fonction du niveau de décision

Lorsqu'on ne prend en compte que le niveau de décision impliqué par un SID, on distingue quatre types de systèmes :

- **EIS "Executive Information System"** (Systèmes d'information pour dirigeants) : Turban (1993) définit l'EIS comme *« un outil fournissant au décideur l'information utile qui lui permet de se focaliser sur les données critiques et d'avoir une bonne appréciation de l'organisation »*. Millet et Mawhinney (1992) considèrent l'EIS comme *« un système qui intègre des*

informations en provenance de sources internes et externes permettant aux dirigeants de contrôler et de demander des informations d'importances vitales pour eux et présentées de façon personnalisée ».

Ce sont des applications de type *tableau de bord* destinées à mettre à la disposition des décideurs tout ou une partie des informations synthétisée et à jour dont ils ont besoin pour mener à bien leur mission.

Un EIS est composé des éléments suivants :

- Une base de données gérée par un SGBD.
- Un ensemble de programmes.
- Un ensemble d'interface de saisie.
- Des écrans d'affichage.
- **ESS "Executive Support System"** (Systèmes d'aide pour dirigeants) : L'ESS va au-delà de l'EIS car il inclue des outils de communication, d'analyse et d'intelligence.

Afin de mieux distinguer les deux systèmes, nous dirons que l'EIS est un outil permettant au décideur de se focaliser sur des données critiques pour obtenir une appréciation de l'organisation. L'ESS doit permettre l'analyse de ces données pour donner une appréciation du futur de l'organisation grâce aux modèles d'analyse. Le passage de l'information à l'intelligence implique de pouvoir affecter le futur (axe temporel et visualisation des liens).

- **DSS "Decision Support System"** (Systèmes d'aide à la décision) : Est un système interactif qui aide le décideur à exploiter les données et les modèles pour trouver une solution à un problème non structuré et analyser l'effet d'éventuels changements de l'environnement sur l'organisation. Le but du DSS est d'aider la décision et non pas de remplacer le décideur. Toutefois, il doit permettre de faire de la planification stratégique, ainsi que de la budgétisation à long terme.
- **PSS "Planning Support System"** (Systèmes d'aide à la planification) : Permet au delà d'un SIAD (DSS), une hiérarchisation et une analyse de la faisabilité des procédures ou décisions admises. Plus précisément, il doit offrir une assistance intelligente en permettant de :

- Formuler les problèmes et les modèles.
- Analyser les problèmes.
- Interpréter et diagnostiquer les situations.
- Formuler des plans d'actions.
- Evaluer des hypothèses.
- Contrôler les résultats des actions par rapport aux attentes.
- Fournir des justifications et des explications.

Si nous revenons au modèle IDC, la phase d'intelligence correspond à l'EIS et l'ESS : le premier permet au décideur de se focaliser sur des données critiques (par exemple la chute des actions en bourse durant une certaine période) et le deuxième de faire une analyse pour une appréciation future de l'organisation. Les autres étapes du modèle IDC de Simon (Simon, 1977) peuvent être assurées selon Turban (1993) par le DSS qui peut faire de la planification stratégique et offre la possibilité d'analyser les effets d'éventuels changements sur l'organisation (en effectuant des simulations) .

2.1.2 Classement en fonction de l'envergure de la décision

Lorsqu'on ne prend en compte que l'envergure de la décision de SID, ils peuvent être classés en trois catégories:

- **Le SID opérationnel:** Il évite la surcharge mentale de l'opérateur en lui proposant des solutions permettant de faire face rapidement à des situations complexes.
- **Le SID de gestion:** Il présente aux responsables opérationnels les indicateurs et alarmes quotidiens utiles au pilotage du travail des opérateurs.

- **Le SID stratégique:** Il présente aux dirigeants des séries chronologiques périodiques éclairant l'efficacité et le positionnement de l'entreprise (résultat, part de marché, tendances).

L'interactivité :

On peut distinguer deux modes de travail en décisionnel :

- **Le mode rapport (non interactif /absence de décideur):** Obtenir une information récurrente, correspond à une analyse prédéfinie.
- **Le mode interactif (présence de décideur):** Chercher une information en effectuant différentes analyses successives, les résultats de l'une amenant des questions nouvelles, qui demandent une nouvelle restitution de données.

2.2. Les qualités indispensables d'un outil d'aide à la décision

En ce qui concerne l'accès aux données, un outil d'aide à la décision doit fournir une interface de requête intuitive simple qui masque la complexité du langage utilisé. Il doit être capable d'adapter les données au contexte de l'utilisateur (une vue métier) en proposant un catalogue pour accéder aux données par sujet.

En ce qui concerne la valorisation des données, l'outil doit être capable de faire des tris, de rajouter des sélections, des totaux...etc. Il doit aussi s'ouvrir aux outils bureautiques ainsi qu'à Internet. L'outil doit s'adapter à plusieurs types d'utilisateurs et doit intégrer le modèle « PULL » et « PUSH » de l'information, c'est-à-dire offrir au décideur la possibilité de demander des informations sur profil et d'être informé lorsque des événements arrivent.

3. Les composants de base d'un système décisionnel

L'architecture des systèmes décisionnels met en jeu quatre éléments essentiels suivants :

- **Les sources de données :** Sont nombreuses, variées, distribuées et autonomes. Elles peuvent être internes (bases de production) ou externes (Internet) à l'entreprise.
- **L'entrepôt de données :** Est le lieu de stockage centralisé des informations utiles pour les décideurs. Il met en commun les données provenant des différentes sources et conserve leurs évolutions.
- **Les magasins de données :** Sont des extraits de l'entrepôt orientés sujet. Les données sont organisées de manière adéquate pour permettre des analyses rapides à des fins de prise de décision.
- **Les outils d'analyse :** Permettent de manipuler les données suivant des axes d'analyses. L'information est visualisée au travers d'interfaces interactives et fonctionnelles dédiées à des décideurs souvent non informaticiens (directeurs, chefs de services,...etc.).

4. Les Systèmes Interactifs d'Aide à la Décision (SIAD)

Les SIAD ont été conçus pour résoudre des problèmes de décision peu ou mal structurée (Keen,1978). Ces problèmes possèdent les ou l'une des propriétés suivantes (Klein, 1971):

- Les préférences, jugements, intuitions et l'expérience du décideur sont essentiels ;
- la recherche d'une solution implique un mélange de recherche d'information, de formalisation ou définition et structuration du problème, du calcul et de la manipulation de données ;
- la séquence des opérations ci-dessus n'est pas connue à l'avance parce qu'elle peut être fonction des données, être modifiée, peut ne donner que des résultats partiels, ou encore peut être fonction des préférences de l'utilisateur ;
- les critères pour la décision sont nombreux, en conflit et fortement dépendant de la perception de l'utilisateur ;
- la solution doit être obtenue en un temps limité ;
- le problème évolue rapidement.

4.1. Définition des SIAD

La littérature regroupe de nombreuses définitions proposées des SIAD (Little 1970, Scott-Morton 1971, Alter 1980, Keen 1980, Moore and Chang 1980) qui ont mis l'accent soit sur le type de problème, soit sur les fonctions du système, soit sur ses composants ou encore sur le processus de développement.

La définition proposée par Bonczek (1981) traite les SIAD (ou encore les DSS) comme STI. Ce travail distingue entre une première partie PPS (Problem Processing System) correspondant au processeur, une deuxième partie KS (Knowledge System) correspondant à des bases de connaissances statiques (mémoires) et une troisième et dernière partie LS (Language System) représentant le moyen de communiquer avec le système. Cette définition des SIAD a l'avantage de souligner le rôle du savoir, mais elle n'est pas assez explicite pour distinguer ce qui fait la spécificité des SIAD. La figure III.1 positionne ces différentes parties.

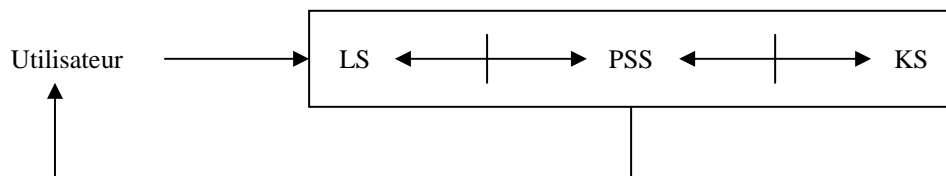


Figure III.1. Principe des SIAD d'après (Bonczek,1981).

On trouve, aussi, cette notion de STI dans la définition de SIAD apportée par Sprague (1982) mais sous une forme un peu différente.

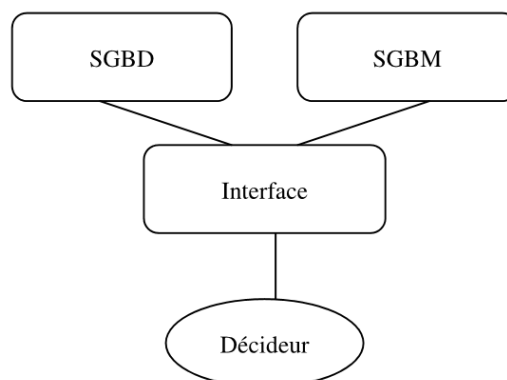


Figure III.2. Principe des SIAD suivant (Sprague,1982).

Cependant, Turban (1993) a proposé une définition des SIAD parlant à la fois des fonctions et de la constitution du système. Il considère qu'un SIAD est un système d'information interactif, flexible, adaptable et spécifiquement développé pour aider à la résolution d'un problème de décision en améliorant la prise de décision. En effet, d'après Turban (1993), un SIAD peut supporter les différentes phases de la prise de décision et inclure une base de connaissances et à travers sa simple interface, sera capable d'utiliser des données comme il autorise l'utilisateur à introduire ou à développer ses propres idées ou points de vue. De plus, il peut être muni d'une base de connaissances comme il peut exploiter des modèles qui peuvent être standards ou spécifiques.

4.2. Caractéristiques des SIAD

Parmi les diverses caractéristiques associées aux SIAD, les suivantes peuvent être rajoutées (Garlatti, 1996) :

- Ils fournissent, principalement, une aide pour les problèmes peu ou mal structurés en considérant, en même temps, l'ensemble des jugements humains et des informations calculées ;
- Ils fournissent une aide personnalisée pour différentes catégories d'utilisateurs, ou des groupes d'utilisateurs ;
- Ils supportent des processus interdépendants ou séquentiels ;
- Ils sont adaptatifs dans le temps en permettant au décideur de réagir face à de nouvelles conditions.
- Ils sont suffisamment flexibles pour que le décideur puisse ajouter, détruire, combiner, changer et réarranger les variables du processus de décision, ainsi que les différents calculs ;
- Comme le SIAD est un système d'aide à la prise de décision et non pas un système de prise de décision, donc, il est conçu pour aider et non pas pour se substituer au décideur. C'est seulement ce dernier qui a le contrôle de toutes les étapes du processus de décision et peut à tout moment remettre en cause les recommandations faites par le SIAD ;
- Ils utilisent des modèles issus des expérimentations de différentes stratégies sous différentes conditions. Ces expériences peuvent apporter de nouvelles vues sur le problème et un apprentissage ;
- Dans le cas de problèmes nécessitant une expertise, des versions avancées de SIAD peuvent être mise en œuvre en exploitant un système à base de connaissances qui apporte notamment une aide efficace et effective.

4.3. Composition d'un SIAD

Comme les SIAD sont des systèmes de résolution de problèmes peu ou mal structurés, donc, ils permettent la recherche heuristique. Pour satisfaire les critères cités ci-dessus, Marakas (2003) a défini des SIAD qui se composent d'une interface homme/machine, d'une base d'informations (de données), d'une base de connaissances et d'une base de modèles (Voir figure III.3).

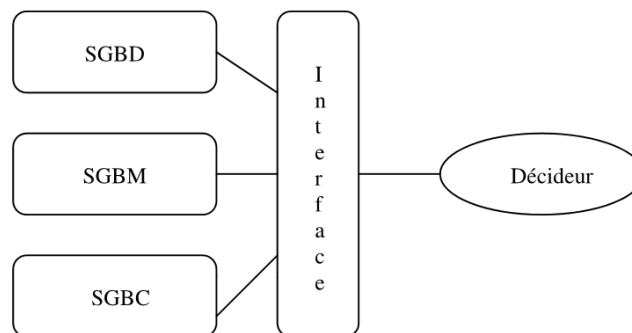


Figure III.3. Composition d'un SIAD. (Marakas,2003)

En premier un SIAD est, pour la tâche qui le concerne, une mémoire de l'entreprise qui l'utilise. Cette mémoire se présente sous la forme d'une base de données. Il faut évidemment pouvoir gérer et exploiter ces données là, ce qui implique qu'elles soient structurées. La réalisation d'un SIAD suppose déjà que les choix des représentations et d'une organisation de données sont faits. Le SIAD fournit donc un cadre pour la représentation de données. Le schéma conceptuel constitue une première représentation de la base de données. Mais le SIAD manipule aussi des représentations plus élaborées comme des graphiques, des courbes, des tableaux, qui sont des représentations synthétiques de données.

L'autre fonction du SIAD est celle qui est liée à la modélisation et au traitement de données. Il s'agit d'effectuer des opérations sur les représentations précédemment définies. Le SIAD fournit donc des moyens de traitement de données et de réalisation de modèles (extraction et agrégation de certaines données, modules de prévision, calculs divers, rentabilité, etc.). On désigne généralement ce traitement sous le terme de *computation*.

Enfin, le SIAD donne à son utilisateur les moyens d'effectuer les tâches précédentes. C'est-à-dire qu'un SIAD comporte des outils destinés à faciliter les manipulations de l'utilisateur dans ses efforts pour arriver à la conclusion de son étude. Ces moyens de manipulation et de contrôle, sont mis à la disposition de l'utilisateur par l'intermédiaire de menus, de masques et de guides.

Les fonctionnalités précédentes peuvent être regroupées en trois sous-systèmes principaux, auxquels il faut ajouter le module de contrôle qui assure la supervision des différentes fonctions :

- Une interface de dialogue qui assure les fonctions de représentation et de manipulation des représentations, c'est aussi par cette interface que passe le contrôle de l'utilisateur.
- Une base de données qui assure la fonction de mémoire.
- Une base de modèles qui assure les fonctions de traitement (computation).

4.4. Types d'architecture des SIAD

Le rôle de l'architecture consiste à réaliser de toutes les façons efficaces possibles, les trois types de modules du SIAD (le dialogue, la base d'informations et les modèles). On distingue en générale quatre types principaux d'architecture : en réseau, centralisée, hiérarchisée et matricielle (Lévine, 1990).

5. Conclusion

Il est possible avec un SID d'alimenter, de stocker et d'exploiter les données à l'aide d'un ensemble d'outils complémentaires disponibles sur le marché décisionnel. Ces outils permettent aux décideurs d'avoir une vision globale sur les activités d'une entreprise par un accès rapide et interactif à un ensemble de vues des données organisées pour refléter l'aspect multidimensionnel des données de l'entreprise.

Les SIAD ont pour objets d'apporter une aide à la décision pour des problèmes peu ou mal structurés. En situation complexe, la décision n'est pas structurée et il devient donc primordial de concevoir des systèmes coopératifs permettant une répartition évolutive des compétences entre l'utilisateur et la machine (dépendante du problème à résoudre) et offrant une bonne intégration des deux agents (homme et machine) dans le processus de décision. Cette répartition est un processus dynamique qui changera au futur et en mesure des problèmes résolus (décisions prises).

6. Références bibliographiques

- Alter, S. L., Decision Support Systems: Current Practices and Continuing Challenges, Alter, S. L., Ed. Addison-Wesley, 1980.
- Bonczek R., Holsapple C. et Whinston A., Foundations of Decision Support System". Academic Press, New York, 1981.
- Garlatti S., Les systèmes interactifs d'aide à la décision en situations complexes. 40èmes Journées de l'APHO (Association de Pharmacie Hospitalière de l'Ouest) : Pharmacien hospitalier et troisième millénaire, Perros-Guirrec, 1996.
- Keen P. et Scott-Morton M., Decision Support Systems: an organizational perspective, Addison-Wesley Publishing, 1978.
- Klein M. and V. Tixier. SCARABEE: a data and model bank for financial engineering and research. IFIP congress, North Holland, 1971.
- Marakas G. Decision Support Systems In the 21 Century. Second Edition, Prentice Hall, 2003.
- Millet I. et Mawhinney C.H., Executive information systems: a critical perspective. Information & Management, Vol 23, num 2, pp 83-92, 1992.
- Moore, J. H. and Chang M.G., Design of Decision Support Systems. Data Base 12 (1 et 2), 1980.

Little, J. D., Models and Managers: The Concept of a Decision Calculus. Management Science 16(8), 1970.

Sprague R. et Carlson E., Building Effective Decision Support Systems. Prentice-Hall, Inc, Englewood Cliffs, 1982.

Turban E., Decision Support and Expert Systems, Macmillan, New York, 1993.

Voir aussi :

Brézillon. P. and Brézillon. J. Context sensitive decision support systems in road safety. Handbook on Decision Support Systems, pp 675-671, Springer-Verlag, International Handbook on Information Systems, 2008.

Carton. F., Adam. F. and Brézillon. P. Why real time transaction processing fails to capture the context required for decision support systems. Supporting real-time decision making : The role of context in decision support on the move, pp 224-234, Annals of Information Systems, Series Decision Support Systems, 2011.

Zararé P. DSS from Theory to Practice. JDS, volume 12, numéros 3 et 4, Hermès-Lavoisier, 2003.

Zararé. P., Des Systèmes Interactifs d'Aide à la Décision Aux Systèmes Coopératifs d'Aide à la Décision : Contributions conceptuelles et fonctionnelles. Habilitation à diriger des recherches, Université Paul Sabatier, Toulouse, 2005.

Chapitre IV : Les arbres de décision

1. Introduction

Les arbres de décision sont des méthodes ou des techniques très connues et utilisées dans le domaine de l'intelligence artificielle, de la fouille de données. Ce sont des méthodes graphiques pour analyser des décisions avec risque, des modèles dans lesquels les probabilités associées aux différents états de la nature sont spécifiées. Plus précisément, les arbres de décisions ont été conçus pour les problèmes faisant intervenir une séquence de décisions et événements successifs. Ce sont une structure qui est souvent utilisée pour représenter les connaissances. En effet ceux-ci formalisent le raisonnement que pourrait avoir un expert désirant retrouver une caractéristique d'un objet.

La construction des arbres de décision à partir de données est une discipline ancienne. Morgan et Sonquist (1963) sont les premiers à avoir utilisé de telles méthodes dans un processus de prédiction et d'explication (AID – Automatic Interaction Detection). Ensuite, Morgan et Messenger (1973) à travers leur travail THAID et Kass (1980) par leur travail CHAID ont arrivé à étendre ces méthodes pour des problèmes de discrimination et classification. Toutefois, c'est la méthode CART (Classification and Regression Tree) de Breiman et al. (1984) qui a fait connaître ces méthodes.

En apprentissage automatique, Quinlan était un chercheur très actif dans la deuxième moitié des années 80. En effet et dans le but d'améliorer le comportement de son système, il a mis en œuvre toute une large gamme de travaux exploitant des heuristiques parmi lesquelles nous pouvons citer ID3 (Induction of Decision Tree) développée en 1979 dont il rattache aux travaux de Hunt (1962). Par la suite, sa méthode C4.5 développée en 1993 était et elle est toujours l'une des références les plus marquantes pour citer les arbres de décision. C'est le même cas pour la méthode C5.0 représentant une évolution de C4.5 dont on ne dispose, malheureusement, pas de détails du fait qu'elle a été implémentée dans un logiciel commercial.

2. Représentation par arbre de décision

Un arbre de décision permet de classer un objet à l'aide de questions : chaque nœud de l'arbre représente une question sur l'un des attributs de l'objet, chaque lien est une réponse à la question (valeur de l'attribut), et chaque feuille est une classe. Ainsi trois types de nœuds sont envisageables :

- ✓ Nœuds de décision : ce sont des nœuds suivis de branches représentant des actions concrètes (contrôlées par le décideur). Ils sont représentés par des carrés.
- ✓ Nœuds événement : ce sont des nœuds suivis de branches représentant les états de la nature. Ils sont contrôlables par le décideur. Ils sont représentés par des cercles. Les événements peuvent être pondérés (attribution de probabilités d'occurrence de l'événement).
- ✓ Les nœuds terminaux (feuilles) : ils correspondent aux résultats attendus.

3. Objectif

Les objectifs à atteindre lors de la conception d'un arbre de décision sont principalement les suivants :

- Être efficace en généralisation (être capable de classer correctement un nouvel exemple).
- Construction du plus petit arbre de décision possible.

4. Construction d'un arbre de décision

La popularité de la méthode repose en grande partie sur sa simplicité. Il s'agit de trouver un partitionnement des individus que l'on représente sous la forme d'un arbre de décision. L'objectif est de produire des groupes d'individus les plus homogènes possibles du point de vue de la variable à prédire. Il est d'usage de représenter la distribution empirique de l'attribut à prédire sur chaque sommet (nœud) de l'arbre.

Pour construire un arbre de décision, nous devons répondre aux 3 questions suivantes :

- Comment choisir, parmi l'ensemble des variables disponibles, la variable de segmentation d'un sommet ?
- Lorsque la variable est continue, comment déterminer le seuil de coupure lors de la segmentation ?
- Comment déterminer la bonne taille de l'arbre ?

Exemple 1:

Supposons la base d'exemples de Ralaivola (2005) décrite par l'ensemble suivant :

| | |
|--------|-------|
| (•, +) | Bleu |
| (•, -) | Rouge |
| (■, +) | Bleu |
| (•, -) | Bleu |
| (◆, -) | Bleu |
| (●, +) | Bleu |
| (◆, -) | Rouge |
| (•, -) | Rouge |

L'ensemble d'attributs qui peut être déduit à travers cette base d'exemples en vue de classer les différentes formes en un ensemble de classes est le suivant :

| Taille | Forme | Couleur | Classe |
|--------|---------|---------|--------|
| Petit | Cercle | Bleu | + |
| Grand | Cercle | Rouge | - |
| Grand | Carré | Bleu | + |
| Petit | Losange | Bleu | - |
| Grand | Losange | Bleu | - |
| Grand | Cercle | Bleu | + |
| Grand | Losange | Rouge | - |
| Petit | Cercle | Rouge | - |

Table IV.1. Base d'exemples 1.

L'arbre de décision modélisant cette base d'exemples est le suivant :

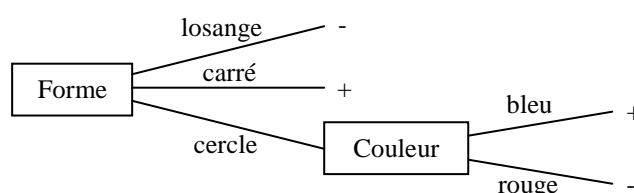


Figure IV.1. Arbre de décision modélisant la base d'exemples 1.

Exemple 2 :

Dans le but de mieux comprendre le déroulement de l'opération de construction d'un arbre de décision, nous allons reprendre et dérouler un exemple qui est présenté dans l'ouvrage de Quinlan (1993). La base de données présentée par la table IV.2, ou encore dite base d'exemples, regroupe 14 exemples décrivant l'influence des prévisions météorologiques sur le comportement des exemples, encore dits individus, par rapport à un jeu {jouer, ne pas jouer}.

| Numéro | Ensoleillement | Température (°F) | Humidité (%) | Vent | Jouer |
|--------|----------------|------------------|--------------|------|-------|
| 1 | soleil | 75 | 70 | oui | oui |
| 2 | soleil | 80 | 90 | oui | non |
| 3 | soleil | 85 | 85 | non | non |
| 4 | soleil | 72 | 95 | non | non |
| 5 | soleil | 69 | 70 | non | oui |
| 6 | couvert | 72 | 90 | oui | oui |
| 7 | couvert | 83 | 78 | non | oui |
| 8 | couvert | 64 | 65 | oui | oui |
| 9 | couvert | 81 | 75 | non | oui |
| 10 | pluie | 71 | 80 | oui | non |
| 11 | pluie | 65 | 70 | oui | non |
| 12 | pluie | 75 | 80 | non | oui |
| 13 | pluie | 68 | 80 | non | oui |
| 14 | pluie | 70 | 96 | non | oui |

Table IV.2. Données "weather" (Quinlan, 1993).

Cette base d'exemple peut être lue comme suit :

Pour pouvoir décider sur le fait de jouer ou non, il suffit de considérer la combinaison des états des quatre attributs "Ensoleillement", "Température", "Humidité" et "Vent". La base d'exemple donnée par le tableau IV.2 présente 14 exemples décrivant cette situation, dont 9 mène à l'état jouer égale à "oui" et 5 à l'état "non".

Donc, pour construire l'arbre de décision correspondant à cette base d'exemples, nous étudions les combinaisons possibles des états ou des valeurs des variables représentatives des quatre attributs cités en haut.

La variable correspondante au premier attribut considéré est appelée *racine* de l'arbre. Prenons l'exemple de l'attribut "ensoleillement", la variable correspondante pourra prendre 3 valeurs possibles {soleil, couvert, pluie}. En terme d'arbre de décision correspondant construit, cela est équivalent à la création de trois branches sortantes à partir de la racine "ensoleillement" représentant un premier niveau de prise de décision menant chacune à un nœud décision de niveau inférieur.

La valeur "Soleil" est recouverte par 5 exemples (de 1 à 5), la valeur "couvert" par 4 exemples (de 6 à 9) et la valeur "Pluie" est recouverte par 5 exemples (de 10 à 14).

Supposons, maintenant, que la première branche à gauche correspond à la valeur "Soleil". Elle est appelée *variable de segmentation*. Parmi les 5 exemples correspondant, on constate qu'il y a 2 exemples « jouer = oui » et 3 « jouer = non ». Cela ne permet pas une prédiction pure, donc ne permet pas une prise de décision unique, ce qui veut dire qu'il faut considérer un autre attribut voire même d'autres attributs pour ce faire (avoir une prédiction unique sur le fait de jouer ou non).

Ce problème ne se présente pas sur la branche correspondante à la valeur "couvert" de l'attribut "ensoleillement" puisque tous les exemples correspondant mènent à l'état "jouer = oui". Donc, le nœud couvert n'aura pas de nœuds fils et il représentera une *feuille* dans l'arbre construit.

Revenons à la première branche et on rappelle qu'il faudra encore considérer plus d'attributs pour chercher une prédiction unique. Supposons que nous allons considérer l'attribut "Humidité".

Comme cette attribut est continu, il sera nécessaire de le discrétiser. Ici, le seuil choisi est 77.5% permettant de créer 2 feuilles correspondantes aux états "humidité >77.5%" et "humidité <77.5%".

On continue ainsi pour le reste de la création de l'arbre de telle manière que sur l'arbre résultant, un individu ne peut pas être placé sur 2 feuilles différentes. Celui modélisant cette base d'exemple est donné par la figure IV.2.

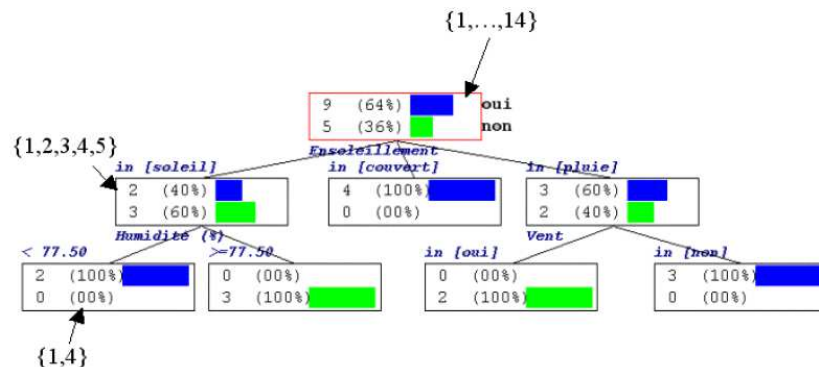


Figure IV.2. Arbre de décision modélisant la base de données "weather". (Rakotomalala, 2005)

L'arbre de décision construit représentant un modèle de prédiction, dans cet exemple, peut être facilement lu. Les branches allant de la racine aux feuilles représentent chacune une alternative qui peut être décrite ou traduite en une *règle de production*, tel que :

- ✓ Les feuilles correspondent à la partie *conclusion* de la règle de production,
- ✓ la combinaison des variables de décision situées aux niveaux supérieurs aux feuilles représente la *condition* de la règle.

Par exemple, d'après l'arbre de la figure IV.2 on peut tirer la règle de production suivante :

Si (Ensoleillement = Soleil) **ET** (Humidité <77.5%) **Alors** Jouer = oui

Toutefois, la création d'un arbre de décision n'est pas aussi facile qu'elle apparaît. Elle couvre de réels difficultés que nous pouvons citer en ce qui suit :

1) Choix de variables de segmentation : Dans l'exemple précédent (exemple 2), nous avons choisi la variable "ensoleillement" comme première variable de segmentation, alors qu'il a été possible de choisir l'une des deux restantes variables ! Le choix doit pouvoir évaluer de manière objective la qualité de la segmentation. Autrement dit, la création des différentes alternatives (la segmentation sur les différents niveaux) doit produire un arbre optimal et en même temps aussi général que possible pour pouvoir classer tout nouvel exemple.

2) Discrétisation : Dans l'exemple 2, lors de l'utilisation de la variable "humidité", il a fallu fixer un seuil (77.5%) pour segmenter la base d'exemples. Quelle est la méthode à appliquer pour ce faire ?

3) Prise de décision unique : Considérons toujours l'exemple 2, notre objectif était de produire un arbre assurant une prédiction pure ou unique, chose qui n'est pas toujours possible ! Nous pouvons penser au fait de développer des règles permettant de définir la taille adéquate de l'arbre de décision. Reste de savoir est ce que c'est faisable ou pas !

4) Choix de modalités dans une feuille : Ce problème ne se pose pas quand la prise de décision est unique sur une feuille, mais plutôt dans le cas où sur une feuille différentes modalités de la variable à prédire se présentent.

La création d'un arbre de décision modélisant une certaine situation doit satisfaire ces quatre points. Il s'agit d'appliquer de méthodes d'induction des arbres de décision à partir de données. Comme en réalité le champ des stratégies possibles étant restreint, donc c'est pas évident de surclasser les techniques existantes en fonction de leur satisfaction des quatre points ci-dessus.

Vers les années 90, plusieurs travaux ont traité de ce point en tentant d'énumérer de manière quasi-exhaustive les variantes sur chacun de ces points. Toutefois, les comparaisons sur données réelles ont montré que ces stratégies produisaient des arbres avec des performances similaires avec des différences notables vis-à-vis leur adaptabilité aux objectifs recherchés (voir par exemple Lerman et Da Costa (1996) pour les descripteurs à très grand nombre de catégories).

La communauté scientifique référence principalement trois méthodes parmi lesquelles des didacticiels sur CART et C4.5 existent en très grand nombre (Nakache et Confais, 2003 ; Kohavi et Quinlan, 2002 ; Bardos, 2001 ; Zighed et Rakotomalala, 2000 ; Lebart et al., 2000 ; Gueguen, 1994 ; Celeux et Lechevallier, 1990).

Dans (Rakotomalala, 2005), l'accent a été mis sur une approche grandement inspirée de la méthode CHAID (CHi-squared Automatic Interaction Detection - Kass, 1980) qui a été l'une des premières méthodes implémentée dans des logiciels commerciaux (SPSS Answer Tree et Knowledge Seeker) et étant particulièrement efficace lorsque la taille de la base de données est importante en utilisant des formulations bien connues en statistique.

5. Création d'un arbre de décision

Dans le but d'explicitier les différentes phases de création d'un arbre de décision optimal, nous allons considérer la même base d'exemples vue précédemment dans l'exemple 2.

5.1 Choisir une variable de segmentation

La plupart des méthodes d'induction d'arbres procèdent comme suit :

pour choisir la variable de segmentation, l'ensemble des exemples est partitionné et un indicateur de qualité est calculé dans le but de déterminer la variable qui optimise cet indicateur. Cette variable sera celle élue pour segmentation. C'est la mesure utilisée qui fait la différence entre une méthode de segmentation et une autre. Plus explicitement, dans une segmentation, un tableau de contingence est défini. Il s'obtient en croisant la variable résultat (variable à prédire ou encore appelée classe) et l'attribut candidat à la segmentation.

Par exemple et en considérant la base d'exemple donnée par le tableau IV.2, et en se plaçant sur la racine et on considérant l'attribut "ensoleillement" candidat à la segmentation, l'application de la méthode expliquée ci-dessus donne le tableau suivant :

| NB Jouer | Ensoleillement | | | |
|----------|----------------|-------|--------|-------|
| Jouer | couvert | pluie | soleil | Total |
| non | 0 | 2 | 3 | 5 |
| oui | 4 | 3 | 2 | 9 |
| Total | 4 | 5 | 5 | 14 |

Table IV.3. Tri croisé à l'aide de la variable "ensoleillement" à la racine de l'arbre. (Rakotomalala, 2005)

La généralisation de cette méthode pour une segmentation en un attribut à L valeurs et une classe (variable à prédire) à k valeurs permet de dresser le tableau illustré par la table IV.4.

| Y/X | x_1 | x_l | x_L | Σ |
|----------|-------|----------------------|-------|----------|
| y_1 | | \vdots | | |
| y_k | | $\dots n_{kl} \dots$ | | n_l |
| y_K | | \vdots | | |
| Σ | | $n_{k.}$ | | n |

Table IV.4. Tableau des effectifs lors du croisement de deux variables. (Rakotomalala, 2005)

CHAID utilise le Khi-2 d'écart à l'indépendance donné par la formule suivante dans le but d'évaluer la pertinence de la variable candidate à la segmentation :

$$\chi^2 = \sum_{k=1}^K \sum_{l=1}^L \frac{\left(n_{kl} - \frac{n_{k.} \times n_{.l}}{n} \right)^2}{\frac{n_{k.} \times n_{.l}}{n}}$$

Comme le critère du Khi-2 varie de 0 à $+\infty$ donc qui avantage les attributs ayant un nombre élevé de valeurs, il est souvent préférable de le normaliser par le nombre de degrés de libertés. Le t de Tschuprow dont le domaine de définition est $[0,1]$ ($t = \frac{\chi^2}{n\sqrt{(k-1)(L-1)}}$) peut être retenu.

Si les attributs comportent le même nombre de valeurs, alors cette variante n'a aucun effet. Toutefois et pour des attributs très disparates, elle semble très efficace.

En reprenant l'exemple abordé ci-dessus (segmentation sur les attributs "ensoleillement" et "vent"), les résultats obtenus suite au calcul du t de Tschuprow sont ceux résumés par le tableau IV.5, où il est clair que l'attribut à retenir est "ensoleillement" ayant la plus grande valeur de t de Tschuprow (0.3559 contre 0.2582 pour l'attribut "vent").

| | T de Tschuprow |
|----------------|----------------|
| Ensoleillement | 0.3559 |
| Vent | 0.2582 |

Table IV.5. Descripteurs discrets candidats sur la racine de l'arbre. (Rakotomalala, 2005)

Ce processus est appliqué sur l'ensemble des nœuds à segmenter. Donc il peut paraître désavantageux en terme de temps d'application ou d'exécution, mais vu que son implémentation est facile à réaliser cela accélère son aboutissement en pratique même sur des bases contenant un grand nombre d'attributs. du fait que sa complexité est linéaire par rapport au nombre d'individus et de variables à condition que la totalité de la base soit stockée dans la mémoire lors de l'exécution. Dans le cas contraire il sera nécessaire de parcourir toute la base sur le disque pour évaluer la pertinence de chaque attribut où cela peut affecter la pertinence recherchée en temps d'exécution.

Catlett (1991) et Chauchat & Rakotomalala (2000) sont des exemples de travaux qui ont tenté d'améliorer la rapidité du système face à de grandes bases de données sans dégrader les performances recherchées.

La littérature énumère beaucoup de travaux cherchant à définir des mesures pertinentes d'évaluation d'un partitionnement dans les arbres de décision. Par exemple, Shih (1999) a proposé une classification suivant un ou plusieurs critères. Wehenkel (1996) a tenté de proposer une formulation générique permettant de retrouver l'ensemble des mesures sous forme de cas particuliers. De leur côté, Mingers (1989) et Buntine & Niblett (1992) ont adopté l'algorithme standard ID3 dans lequel la mesure à tester est substituée à l'indicateur originel représentant le gain d'entropie de Shannon pour tester leurs performances. Ils arrivent tout les deux à la conclusion dénonçant que dès lors que les mesures utilisées possèdent de bonnes propriétés de spécialisation, c'est-à-dire tendent à mettre en avant les partitions avec des feuilles pures (prédiction unique), elles ne jouent pas un rôle majeur dans la qualité de la prédiction. Breiman *et al.*, promoteurs de la méthode CART, ont d'ailleurs arrivé à la même conclusion en 1984.

Un autre point important qu'on peut le signaler, ici, est celui de l'instabilité des arbres de décision qui se traduit par le fait que les choix effectués sur les parties hautes de l'arbre ne sont pas sans conséquence sur les choix réalisés sur les parties basses. Ceci se produit dans le cas où les attributs possèdent un pouvoir prédictif équivalent ce qui rend la détection de la variable correspondant au maximum fortement dépendante de l'échantillon d'apprentissage. Dans ce cas, la

moindre modification dans l'échantillon d'apprentissage entraîne la génération d'un arbre de décision visuellement très différent.

Les praticiens comparent cette caractéristique d'instabilité à des méthodes linéaires, comme l'analyse discriminante, où des modifications mineures dans l'échantillon se traduisent par une variation faible des coefficients calculés. Mais en réalité cette variabilité de prédiction n'est pas aussi forte du fait qu'on attribut, généralement, la même étiquette pour un exemple pris au hasard dans l'échantillon.

5.2. Traiter les variables continues

Considérons, maintenant, l'attribut "humidité" placé comme deuxième niveau de prise de décision partant de la racine "ensoleillement" et recouvrant les 5 premiers exemples de la base de l'exemple2. Le seuil de coupure utilisé étant « 77.5 % » a été obtenu comme suit (Rakotomalala, 2005) :

- 1) Sélectionner la meilleure valeur de coupure pour chaque variable continue ;
- 2) Sélectionner globalement la meilleure segmentation en comparant la pertinence de tous les attributs : les attributs discrets et les attributs continus qui ont été découpés en 2 intervalles.

a. Choisir un point de coupure

La première opération consiste à sélectionner la meilleure valeur de coupure pour chaque variable continue. Dans (Rakotomalala, 2005), l'auteur considère le cas du découpage binaire mais il est possible de modifier cette discrétisation avec une valeur seuil différente pour des nœuds situés dans des niveaux plus bas dans l'arbre. Notons qu'aucun avantage mis à part la réduction visuellement du nombre de niveaux de l'arbre, sans en réduire le nombre de feuilles, n'a été signalé par les études cherchant à évaluer l'opportunité d'une discrétisation n-aire.

De plus, le choix du seuil de discrétisation doit dépendre de la procédure de sélection des variables de segmentation ce qui justifie l'application du t de Tschuprow servant à évaluer les partitions dans le choix de la borne.

Résumons, finalement, ce procédé de discrétisation d'un attribut continu X comme suit :

- ✓ Trier les données selon les valeurs de X ,
- ✓ Calculer le t de Tschuprow du tableau de contingence formé temporairement pour tester chaque borne de coupure possible entre deux valeurs de la variable.

La figure IV.3 présente le cas de discrétisation de la variable "humidité" possédant 4 valeurs différentes {70, 85, 90, 95} pour les 5 premières observations correspondantes à la valeur "soleil" du nœud "ensoleillement". Le point de coupure est, généralement, pris à mi-chemin entre 2 points successifs mais, en réalité, toute valeur située dans l'intervalle pourrait être utilisée. Ensuite le tableau de contingence rassemblant les valeurs de l'indicateur associé (t de Tschuprow pour le cas de (Rakotomalala, 2005)) est formé. Nous constatons que le meilleur découpage produit une partition unique, avec un Tschuprow égal à 1. La borne de découpage optimale est 77.5 %.

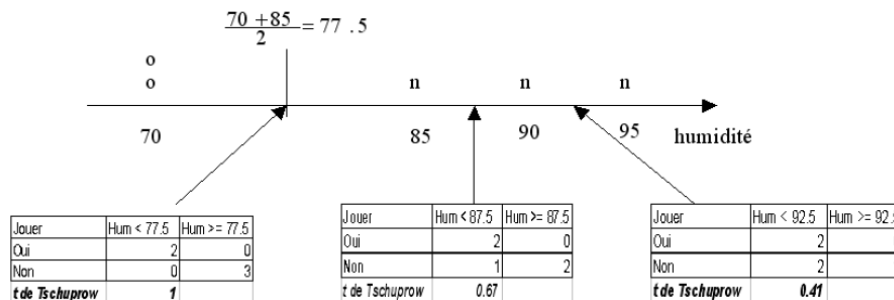


Figure IV.3. Sélection de la borne de discrétisation. (Rakotomalala, 2005)

De façon générale, la discrétisation se réalise sur deux étapes dont la première consiste à trier les données, et la deuxième consiste à tester chaque point de coupure candidat et retenir celui qui optimise l'indicateur de qualité du partitionnement choisi.

Pour des bases d'exemples de tailles raisonnables, le temps d'exécution est acceptable, mais quand la taille de la base devienne critique en englobant des centaines de milliers d'exemples avec un grand nombre d'attributs continus, la création de l'arbre de décision va se tarder beaucoup plus sur le tri de données et le test des points de coupure. Plusieurs stratégies ont été proposées pour améliorer le temps d'exécution dans ce dernier cas (grosses bases d'exemples). Witten et Franck (2000) ont proposé d'ordonner les données une fois pour toute avant de lancer l'exécution en conservant par la suite un index des valeurs triées. L'optimisation proposée par Chauchat et Rakotomalala (2000) dénonce la possibilité de ne traiter qu'un échantillon réduit d'exemples présents sur le nœud sans que cela dégrade la qualité de l'apprentissage. Fayyad et Irani (1993) et Muhlenbach et Rakotomalala (2005) dans leurs travaux ont montré l'impossibilité d'amélioration de l'indicateur de qualité de partition avec un point de coupure situé entre deux exemples de même étiquette. C'est pourquoi ça ne serait pas nécessaire de tester les points de coupures situés entre deux exemples portant la même étiquette (les deux bornes 87.5 et 92.5 de l'exemple précédent).

b. Sélectionner la variable de segmentation

L'étape qui suit la détermination du point de coupure optimal pour chacune des variables continues expliquée dans la section précédente est celle qui consiste à déterminer la variable de segmentation pour le nœud traité. C'est une étape assez simple qui consiste à sélectionner parmi l'ensemble des variables, discrètes ou continues discrétisées, celle qui maximise la mesure de référence sur le nœud en cours de traitement (t de Tschuprow pour (Rakotomalala, 2005)).

En considérant toujours l'exemple 2 et on poursuivant la résolution à partir des résultats obtenus dans l'étape précédente (a), la table IV.6 résume les valeurs calculées de t de Tschuprow. C'est clair que "humidité" est la variable optimale, chose non étrange du fait qu'elle permet des prédictions uniques.

| Descripteur | Point de coupure | T de Tschuprow |
|-------------|------------------|----------------|
| Humidité | 77.5 | 1.00 |
| Température | 77.5 | 0.67 |
| Vent | - | 0.17 |
| Soleil | - | 0.00 |

Table IV.6. Segmentations candidates et bornes de discrétisation associées pour les descripteurs continus. (Rakotomalala, 2005)

Comme le calcul de la borne de discrétisation recherchée localement lors de la segmentation est fortement dépendant de l'échantillon d'apprentissage (base d'exemples), donc elle peut être très instable et la valeur obtenue peut ne pas être interprétable pour l'expert du domaine. Pour remédier à ce problème, plusieurs stratégies ont été proposées.

Une première stratégie intuitive consiste en celle donnant la main à l'expert d'intervenir dès qu'un résultat sera proposé par l'algorithme de calcul adopté et de faire introduire une valeur de coupure plus appropriée pour un nœud, alors que le reste de l'arbre peut être construit automatiquement. Suarez et Lutsko (1999) ont proposé un processus de décision moins immédiat qui consiste à définir une distribution de points de coupures au lieu d'une estimation ponctuelle. Il s'agit de définir un point de coupure *flou*, chose qui va nuire à la lecture de l'arbre du fait qu'on se plaçant sur un nœud, on sera redirigé sur plusieurs feuilles avec des poids différents. Cependant, cette stratégie réduit considérablement la variabilité des arbres de décision. Dans leur travail, Dougherty et al. (1995) distinguent entre deux phases : une première phase de discrétisation dite de prétraitement et une deuxième de construction de l'arbre sur des données pré-discrétisées. Ils ont montré que les performances d'une discrétisation locale sont très proches de celle d'une discrétisation globale des variables.

5.3. Définir la bonne taille de l'arbre

D'après Breiman *et al.* (1984), la taille de l'arbre représente un point crucial pour déterminer la performance d'un arbre de décision permettant de produire un « classifieur » trop complexe. Et d'après Oates et Jensen (1997), cette taille a tendance à croître avec le nombre d'exemples dans la base d'apprentissage ou encore dite base d'exemples.

La figure IV.4 présentant l'évolution du taux d'erreur en apprentissage et en test démontre la nécessité de déterminer une règle suffisamment efficace pour assurer les meilleures performances à l'arbre de décision. En effet, on constate que lorsque le nombre de feuilles, représentant la taille de l'arbre, augmente, le taux d'erreur calculé sur les données d'apprentissage diminue constamment. Quand au taux d'erreur calculé sur l'échantillon test, il montre d'abord une décroissance rapide, jusqu'à un arbre comportant une quinzaine de feuilles, puis il reste sur un plateau avant de se dégrader lorsque l'arbre est manifestement surdimensionné.

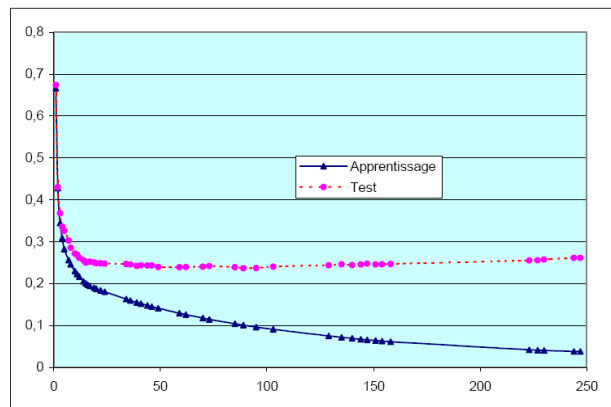


Figure IV.4. Variation du taux d'erreur en apprentissage et en test. (Rakotomalala, 2005)

Quand le taux d'erreur commence à stagner, la détermination de la taille optimale de l'arbre cherche à stopper (opération appelée *pré-élagage*) ou à réduire (opération appelée *post-élagage*) l'arbre dans le but d'obtenir un classifieur correspondant au *coude* de la courbe sur échantillon test. Les deux sections qui suivent décrivent les deux opérations : pré-élagage implémentée par CHAID et le post-élagage.

a. Pré-élagage

Le pré-élagage consiste à stopper la construction de l'arbre en fixant un critère d'arrêt. On pourra, par exemple, fixer un critère d'arrêt local relatif au nœud en cours de traitement permettant d'évaluer *l'apport informationnel* de la segmentation à initier.

Le critère d'arrêt adopté par la méthode CHAID permet de stopper la construction de l'arbre sur un nœud, ce qui revient à accepter la segmentation d'un nœud, si le Khi-2 calculé (ou le t de Tschuprow) sur ce nœud est significativement supérieur à un seuil théorique fixé correspondant au risque de première espèce. Ceci est équivalent au fait de stopper la construction quand la p -value calculée est inférieure au seuil fixé. De ce fait CHAID a été reconnue comme méthode de pré-élagage cohérente.

Sur le même exemple précédemment considéré, la variable "humidité" a été utilisée pour segmenter le nœud racine "ensoleillement". Le t de Tschuprow correspondant égal à 1.0. En adoptant le Khi-2 comme critère d'indépendance, la p -value calculée sera égale à 0.025. Dans ce cas, pour un seuil fixé égale à 5%, la segmentation sera acceptée. Pour un seuil fixé à 1%, elle sera refusée. On conclut que l'arbre construit est fortement dépendant de la valeur du risque choisie. Donc, on doit fixer le bon seuil pour avoir une bonne segmentation, chose qui n'est pas facile.

En effet, il est très difficile de choisir correctement le seuil dans la pratique. Dans le cas où il est trop restrictif, l'arbre sera sous-dimensionné. Dans le cas contraire, l'arbre sera sur-dimensionné.

Pour la valeur du seuil qui égale à 1%, précédemment testée, si elle été adoptée, l'arbre aurait été stoppé dès la racine !

Et comme pour un problème de prévision, ce critère d'arrêt (fixer une valeur de risque) n'a aucun lien direct avec l'objectif de construire un arbre de décision le plus précis possible, donc ce problème est théoriquement insoluble. En effet, la mesure symétrique Khi-2 représentant un test d'indépendance statistique ne reflète pas une situation de prévision. Donc, il n'est plus approprié pour l'utiliser comme test dans une telle situation. De plus et dans le but de contrôler la taille de l'arbre, on fixe un seuil très bas quand les effectifs sont élevés. Il est utile de noter, aussi, que le comportement global de l'arbre n'est pas pris en considération du fait que l'évaluation est locale à un nœud.

Cependant et malgré tous les points limites cités ci-dessus, cette approche donne de bons résultats. La figure IV.4 démontre cela, la plage dans laquelle l'erreur est faible est relativement large, il suffit donc de proposer une règle adéquate pour obtenir un arbre convenable à condition, bien sûr, qu'il ne soit pas sous-dimensionné. Dans le même sens, Jensen et Cohen (2000) défendaient cette approche en disant que le test d'indépendance utilisé n'est pas classique car la variable testée a été produite sur plusieurs étapes d'optimisation partant de la recherche du point de discrétisation optimal pour les variables continues suivie de la recherche de la variable de segmentation qui maximise la mesure utilisée. Donc, les comparaisons sont multiples et la loi statistique n'est plus la même.

b. Post-élagage

C'est avec la méthode CART de Breiman *et al* (1984) que cette approche a vu le jour, pour être, ensuite, très largement reprise sous différentes formes. Elle consiste à appliquer deux phases dans le but de construire un arbre de décision. La première phase est celle d'expansion sur laquelle toutes les segmentations seront acceptées même si elles ne sont pas pertinentes, l'essentiel c'est de produire des arbres les plus purs possibles. Ce principe est aussi appelé *hurdling*. La deuxième phase est celle de réduction de l'arbre en comparant des arbres de tailles différentes. Son intérêt majeur est la construction d'arbres plus performants en classement malgré leur temps d'exécution qui soit élevé surtout pour les grosses bases d'exemples.

Les approches de post-élagage se distinguent en deux catégories différentes. Une première qui cherche à transformer le problème d'apprentissage en un problème d'optimisation par exploitation de transformations bayésiennes ou des dérivées telles que la théorie de la description minimale des messages. Wallace et Patrick (1993) disaient que le critère adopté par cette première catégorie d'approches traduit le compromis entre la complexité de l'arbre et son aptitude à coller aux données ou celui établissant un compromis entre la quantité d'informations nécessaire pour décrire l'arbre, et les données qui font exception à l'arbre pour la théorie de la longueur minimale des messages. Cependant, ces méthodes sont peu connues. D'ailleurs elles ne sont implémentées que dans quelques programmes distribués sous forme de code source (Buntine, 1991 ; Kohavi et Sommerfield, 2002).

La deuxième catégorie d'approches traite, lors de la phase d'élagage, des estimations non-biaisées du taux d'erreur en classement. Plusieurs travaux s'inscrivant sous cette catégorie ont été réalisés. Celui de Quinlan (1993) utilise une estimation calculée sur le même échantillon d'apprentissage mais pénalisée par la taille de l'effectif du nœud à traiter. CART de Breiman *et al*. (1984) utilise une évaluation du taux d'erreur avec un deuxième échantillon, dit de validation. Ces deux travaux ont été étudiés par Kohavi et Quinlan (2002) pour arriver à la conclusion dénonçant l'adaptabilité du premier travail pour des fins d'apprentissage automatique et du deuxième pour des fins statistiques. Cependant dans la pratique, c'est CART qui se révèle plus robuste car elle englobe tous les éléments nécessaires pour un apprentissage efficace, à savoir :

- ✓ Pour déterminer le bon arbre, l'évaluation non biaisée de l'erreur est utilisée,

- ✓ Risque limité de sur-apprentissage sur l'échantillon de validation grâce à la réduction de l'espace des hypothèses avec le principe des séquences d'arbres rangés à coût-complexité décroissant,
- ✓ préférence donnée à la simplicité avec la règle de « l'écart-type » avant l'erreur minimale : l'idée est de se rapprocher du coude dans l'évolution de l'erreur en fonction du nombre de feuilles de l'arbre (Figure IV.4).
- ✓ Utilisation d'un système de validation croisée pour réaliser le post-élagage de fichiers d'apprentissage de tailles réduites.

5.4. Décider

La dernière étape de la construction de l'arbre est celle de décider. Elle permet d'affecter une conclusion à chaque feuille de l'arbre. Ainsi, le chemin reliant une feuille à la racine de l'arbre peut être lu comme une règle de prédiction du type attribut-valeur « Si prémisse alors Conclusion ».

Deux cas se présentent, cas de feuilles pures (feuilles à modalité unique) et celui de feuilles présentant plusieurs modalités. Dans le premier cas, la conclusion correspond à l'unique modalité (valeur de l'attribut) présente. En reprenant l'exemple précédemment considéré et comme toutes les feuilles étant pures, les 5 règles déduites sont celles résumées par la table IV.7. Dans le deuxième cas, la règle d'attribution la plus souvent utilisée est la règle de la majorité qui consiste à affecter à la feuille la modalité de la variable à prédire présentant le plus grand effectif. Cette règle repose sur des fondements théoriques bien établis. En effet, la distribution de fréquences visible sur la feuille est une estimation de la probabilité conditionnelle d'appartenance à chaque étiquette de la variable à prédire. Donc, affecter à la feuille l'étiquette la mieux représentée minimise la probabilité de mauvaise affectation sous deux conditions : (1) les données constituent un échantillon représentatif de la population ; (2) les coûts de mauvaise affectation sont unitaires où les bonnes affectations coûtent 0, et les mauvaises affectations coûtent 1.

| N° | Prémisse | Conclusion |
|----|---|-----------------|
| 1 | Ensoleillement = « Soleil » ET Humidité < 77.5 | Jouer = « oui » |
| 2 | Ensoleillement = « Soleil » ET Humidité >= 77.5 | Jouer = « non » |
| 3 | Ensoleillement = « Couvert » | Jouer = « oui » |
| 4 | Ensoleillement = « pluie » ET Vent = « oui » | Jouer = « non » |
| 5 | Ensoleillement = « pluie » ET Vent = « non » | Jouer = « oui » |

Table IV.7. Règles extraites de l'arbre de la Figure IV.2. (Rakotomalala, 2005)

Bardos (2001), dans son ouvrage, présente un exemple détaillé démontrant inadaptabilité de la règle de la majorité dans la plupart des études réelles où les coûts de mauvaise affectation ne sont pas symétriques. Il disait que la conclusion devrait être celle qui minimise le coût moyen de mauvaise affectation.

5.5. Fusionner les sommets lors de la segmentation

La fusion des sommets lors de la segmentation rentre dans le cadre d'optimisation de l'arbre à construire. Nous pouvons citer, à ce stade, trois travaux ayant différents objectifs vis-à-vis l'optimisation. Le premier travail est celui de Kass (1980) donnant lieu à la méthode CHAID ayant pour objectif la fusion de nœuds enfants issus d'une segmentation. Le deuxième travail est celui de Breiman *et al.* (1984) inventant la méthode CART. Cette dernière impose l'induction d'arbres binaires où les valeurs d'attributs (appelées aussi modalités) sont regroupées en deux sous-ensembles sans justificatifs dans le but d'optimiser l'indicateur de qualité de la partition. Son objectif majeur étant plutôt l'optimisation du regroupement binaire effectué en réduisant le nombre de calculs nécessaires pour le réaliser. Le troisième travail est celui de Quinlan (1993). Il cherchait à optimiser de manière empirique les regroupements à effectuer pour conclure, finalement, au fait que

ce mécanisme permet de réduire la *largeur* de l'arbre sans améliorations notables relatives aux performances de classement.

Ensuite, Rakotomalala (1997) a venu pour confirmer les avantages de tels regroupements à savoir :

- ✓ L'amélioration de la lisibilité de l'arbre en isolant les modalités non-informatives des attributs en luttant contre la fragmentation, surtout, préjudiciable lorsque l'on travaille sur des petits effectifs,
- ✓ réduction de la taille de l'arbre en évitant que des séquences de segmentations se répètent dans différentes zones de l'arbre (la « réplication des sous-arbres »)

Prenons l'exemple de la méthode CHAID, nous décrivons ci-dessous son mécanisme proposé de fusion ou de regroupement de sommets qui se résume au fait de fusionner itérativement, sur un voisinage des profils des sommets enfants issus de la segmentation, les sommets produisant des feuilles avec des distributions similaires. Cette méthode utilise le test d'équivalence de distribution du Khi-2 où le seuil de coupure (risque de première espèce du test) est fixé par l'utilisateur. Il est à noter que la statistique du Khi-2 suit une loi du χ^2 à (K-1) degrés de liberté sous l'hypothèse d'égalité des distributions, tel que :

$$\chi^2 = \sum_{k=1}^K \frac{\left(\frac{n_{k1}}{n_1} - \frac{n_{k2}}{n_2} \right)^2}{\frac{n_{k1} + n_{k2}}{n_1 \times n_2}}$$

Autrement dit, cette méthode cherche à fusionner, tout d'abord, deux à deux les feuilles à profils proches. Ensuite, cette opération sera réitérée sur les feuilles restantes jusqu'à ce qu'aucune fusion ne soit plus possible.

Remarque :

Lors d'une segmentation, deux cas particuliers peuvent être rencontrés :

- 1) Il se peut qu'il n'y ait aucune fusion à réaliser ;
- 2) il se peut que tous les sommets enfants soient fusionnés dans un seul groupe. Dans ce cas, la segmentation avec l'attribut considéré sera rejetée.

Exemple :

Déroulons la démarche de CHAID sur le même exemple considéré dans les sections précédentes (Exemple 2). Etant donné que la racine de l'arbre soit l'attribut "ensoleillement", donc trois modalités candidates à la fusion se présentent : "soleil" notée par "a" pour faciliter la lecture, "couvert" notée par "b" et "pluie" notée par "c". Soit un risque de première espèce du test d'équivalence distributionnelle égale à 10%.

Rappelons que la première passe de CHAID essaie de fusionner les sommets deux à deux. Les résultats des calculs sont résumés sur la Table IV.8.

| Sommets | Distribution | CHI-2 | p-value | Sortie |
|---------|--------------------|-------|---------|---------------|
| a & b | (2 ; 3) et (4 ; 0) | 3.60 | 0.058 | |
| a & c | (2 ; 3) et (3 ; 2) | 0.40 | 0.527 | fusion |
| b & c | (4 ; 0) et (3 ; 2) | 2.06 | 0.151 | |

Table IV.8. 1ère passe, fusion deux à deux des sommets pour la segmentation de la racine "ensoleillement". (Rakotomalala, 2005)

Le contenu de la table parle de lui même. Les sommets (a & c) et (b & c) montrent une p-value supérieure au seuil fixé à 10%, donc, elles peuvent être fusionnées. Et comme ceux sont les distributions des sommets (a & c) qui sont les plus proches, donc, ils seront fusionnés.

Supposons que nous allons re-noter les sommets fusionnés (a & c) en (A) et le sommet (b) en (B). La table IV.9 résume les calculs effectués sur la deuxième passe, d'après lesquels, il est clair qu'aucune fusion n'est possible car les deux sommets présentant des distributions différentes.

| Sommets | Distribution | CHI-2 | p-value | Sortie |
|---------|--------------------|-------|---------|--------|
| A & B | (5 ; 5) et (4 ; 0) | 3.11 | 0.078 | |

Table IV.9. 2ème passe, fusion des sommets pour la segmentation de la racine "ensoleillement". (Rakotomalala, 2005)

La figure IV.5 présente une modélisation de l'opération de segmentation de la racine de l'arbre, "ensoleillement", avec fusion des sommets enfants.

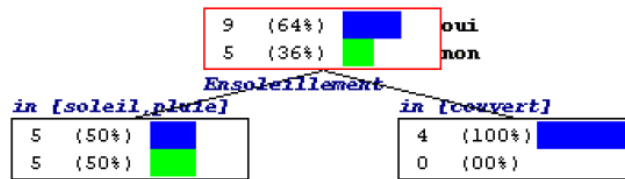


Figure IV.5. Segmentation de la racine de l'arbre avec fusion des sommets enfants. (Rakotomalala, 2005)

Remarque :

Vu que la segmentation est forcément binaire dans le cas de descripteurs continus, alors, cette technique ne joue aucun rôle dans ce cas.

6. Avantages et faiblesses des arbres de décision

6.1. Avantages

Cette technique permet de systématiser l'étude d'un problème de décision complexe en incertitude. Elle est autant populaire en statistique qu'en apprentissage automatique. Son succès réside en grande partie à ses caractéristiques :

- **lisibilité** du modèle de prédiction, l'arbre de décision, fourni. Cette caractéristique est très importante, car le travail de l'analyste consiste aussi à faire comprendre ses résultats afin d'emporter l'adhésion des décideurs.
- Nombre de tests limités par le nombre d'attributs (de question).
- Construction efficace (mais technique) à l'aide d'apprentissage par optimisation (pour obtenir un arbre petit et « correct »).

6.2. Faiblesses

Les faiblesses et limites des arbres de décision peuvent se résumer en ce qui suit :

- Bien que les arbres de décision soient riches en informations, leur utilisation peut faire face à une explosion du nombre d'états de la nature et d'action.
- Ils ne permettent pas de modéliser les rapports de causalité qui peuvent lier les nœuds entre eux.
- Caractère subjectif des probabilités utilisées
- Évaluation des coûts et profits associés à chaque cas de figure
- En cas de risque important : décès, handicap, ruine du système... l'espérance mathématique n'est pas aussi représentative car il faut pouvoir supporter la perte :

Événement A : $P(A) = 0,85$

Profit = 5 000

Événement B : $P(\text{Non}(A)) = 0,15$

Profit = - 20 000

$E(x) = 1250$

Mais la perte de 20 000 est elle supportable ?

7. Evaluation d'un arbre de décision

7.1. Les critères de décision non probabilistes

Ils correspondent à une prise de décision sans probabilité (problèmes équiprobables). Dans ce cas, les critères d'évaluation sont les suivants (Espinasse, 2009) :

- **Critère Maximax** : « le critère du décideur optimiste » : Dans ce cas, le décideur se soucie du gain. Il essaie de choisir la meilleure solution selon le critère « maximiser le gain ». Ce critère est appelé **Maximax**. Pour chaque décision, il choisisse comme solution le meilleur cas.
- **Critère Maximin (Critère de Wald)** : « le critère du décideur pessimiste » : Dans ce cas le décideur se soucie beaucoup plus du risque de la perte. Il essaie, alors, de choisir la meilleure solution selon le critère « minimiser la perte ». Ce critère est appelé **Maximin**. Il consiste pour chaque décision à chercher le plus mauvais cas.
- **Restriction de la perte** : Ce cas présente un optimisme limité, dans le sens où le décideur espère maximiser le gain mais avec un risque limité ; c'est à dire, il cherche le meilleur gain avec un risque de perte ne dépassant pas un certain seuil.
- **Regret Minimax (Critère de Savage)** : Un regret correspond à la différence entre le gain maximal et le gain réel pour chaque colonne du tableau présentant l'utilité de chaque alternative. Dans ce cas, le décideur optera pour la minimisation du regret maximal. On appelle ce critère le « le Regret Minimax ». Pour chaque ligne (décision possible), on cherche le regret maximal, puis on choisit la ligne ayant le plus petit regret maximal.
- **Critère de Laplace** : Pour chaque alternative, on calcule le gain moyen puis on choisit l'alternative ayant le maximum de gain moyen.

Exemple 2 :

Cet exemple traite les investissements immobiliers. Il y a lieu de s'investir dans :

- une résidence
- un immeuble
- un appartement
- aucun investissement

Cela va dépendre de l'état du marché immobilier : fort, moyen ou faible.

Suivant la décision et l'état du marché, le profil dans chacun des cas est donné à travers le tableau suivant :

| | Fort | moyen | faible |
|-------------|------|-------|--------|
| Résidence | 550 | 110 | -310 |
| Immeuble | 300 | 129 | -100 |
| Appartement | 200 | 100 | -32 |
| Aucun | 0 | 0 | 0 |

Table IV.10. Base d'exemples : Investissement immobilier.

- 1) Modéliser la situation ci-dessus selon le principe de l'arbre de décision.
- 2) Suivant les différents critères d'évaluation, pour quelle solution le décideur optera dans chacun des cas ?

7.2. Décision sous risque

Le problème tel que décrit dans l'exemple précédent correspond à une analyse de prise de décision sous incertitude. Si le décideur dispose d'informations sur les probabilités de chaque situation, on appelle ce problème un problème de décision avec risque. L'analyse des décisions dans ce cas est une généralisation du critère de Laplace (qui considère les probabilités égales pour chaque état de la nature). Dans ce cas, au lieu de calculer la moyenne des gains, on calcule l'espérance des gains (respectivement, l'espérance des regrets) et on choisit l'alternative qui maximise la première (respectivement, qui minimise l'espérance des regrets).

L'espérance de gain est donnée par la formule : $E(G_j) = \sum_{i=1}^n P_i G_{ij}$

L'espérance des regrets est donnée par la formule : $E(R_j) = \sum_{i=1}^n P_i R_{ij}$

Où :

i est l'état de la nature (une situation ou un cas),

P_i est la probabilité associée à l'état i ,

G_{ij} et R_{ij} sont respectivement le gain et le regret associé à cet état i

Cette manière de calculer s'appelle **critère de Bayes**.

Exemple 3 :

Nous considérons la même situation présentée à travers l'exemple 3, mais on dispose de plus, cette fois ci, des probabilités suivantes :

Résidence → 10%

Immeuble → 40%

Appartement → 30%

Aucun → 20%

Question : Calculer l'espérance de gain et celle de regret, et dites ensuite pour quelle solution optera le décideur.

8. Références bibliographiques

Bardos M., *Analyse Discriminante : Application au risque et scoring financier*, Dunod, 2001.

Breiman L, Friedman J. et Olshen R., Stone C., *Classification and Regression Tree*, California: Wadsworth International, 1984.

Buntine W., *About the IND tree package*, Technical Report, NASA Ames Research Center, Moffet Field, California, September 1991.

Buntine W., Niblett T., *A further comparison of splitting rules for decision tree induction*, Machine Learning, 8, pp 75-85, 1992.

Catlett J., *Megainduction : machine learning on very large databases*, PhD Thesis, University of Sidney, 1991.

Celeux G., Lechevallier Y., *Méthodes de segmentation*, in *Analyse Discriminante sur Variables Continues*, Celeux G. éditeur, INRIA, 7, pp 127-147, 1990.

Chauchat J.H., Rakotomalala R., *Sampling Strategy for Building Decision Trees from Very Large Databases Comprising Many Continuous Attributes*, in *Instance Selection and Construction for Data Mining*, Liu H. and Motoda H. Editors, Kluwer Academic Press, pp 171-188, 2000.

Dougherty J, Kohavi R., Sahami M., *Supervised and unsupervised discretization of continuous attributes*, in *Proceedings of 12th International Conference on Machine Learning*, pp 194-2002, 1995.

Espinasse B., *Analyse de la décision dans l'Incertain*. Cours donné à l'Université d'Aix-Marseille, 2009. <http://www.lsis.org/espinasseb/Supports/MOAD-2009/4-AnalyseDecisionIncertain-2009-4P.pdf>

Fayyad U, Irani K., *Multi-interval discretization of continuous attributes for classification learning*, in *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pp 1022-1027, 1993.

- Gueguen A., *Arbres de décision binaires*, in Analyse Discriminante sur Variables Qualitatives, G. Celeux et J.P. Nakache Editeurs, chapitre 7, Polytechnica, 1994.
- Hunt E.B., *Concept Learning: An Information Processing Problem*, Wiley, 1962.
- Jensen D., Cohen P., Multiple Comparisons in Induction Algorithms, *Machine Learning*, 38(3), pp 309-338, 2000.
- Kass G., An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, 29(2), pp 119-127, 1980.
- Kohavi R., Quinlan J., Decision-tree Discovery, in *Handbook of Data Mining and Knowledge Discovery*, Klösgen and Zytkow Editors, pp 267-276, 2002.
- Kohavi R., Sommerfield D., MLC++. In Will Klossgen and Jan M. Zytkow, editors, *Handbook of Data Mining and Knowledge Discovery*, chapter 24.1.2, pp 548-553. Oxford University Press, 2002.
- Lebart L., Morineau A., Piron M., *Statistique exploratoire multidimensionnelle*, Dunod, 2000.
- Lerman I., Da Costa F., *Coefficients d'association et variables à très grand nombre de catégories dans les arbres de décision, application à l'identification de la structure secondaire de protéines*, Rapport INRIA, n°2803, Février 1996.
- Mingers J., *An empirical comparison of selection measures for decision tree induction*, *Machine Learning*, 3, pp 319-342, 1989.
- Morgan J., Messenger R., THAID-a sequential analysis program for the analysis of nominal scale dependent variables, Survey Research Center, U of Michigan, 1973.
- Morgan J., Sonquist J.A., Problems in the Analysis of Survey Data, and a Proposal, *Journal of the American Statistical Association*, 58:415-435, 1963.
- Muhlenbach F., Rakotomalala R., *Discretization of Continuous Attributes*, Encyclopedia of Data Warehousing and Mining, Wang J. editor, Idea Group Reference, 2005.
- Nakache J-P., Confais J., *Statistique Explicative Appliquée*, Edition Tecnip, Paris, 2003.
- Oates T., Jensen D., The effects of Training Set Size on Decision Tree Complexity, in *Proceedings of 14th International Conference on Machine Learning*, pp 254-262, 1997.
- Quinlan R., Discovering rules by induction from large collections of examples, D. Michie ed., *Expert Systems in the Microelectronic age*, pp 168-201, 1979.
- Quinlan R., *C4.5: Programs for Machine Learning*, Morgan Kaufman, 1993.
- Rakotomalala R., *Graphes d'Induction*, PhD Thesis, Université Claude Bernard Lyon 1, 1997.
- Rakotomalala R., Arbres de décision. *Revue MODULAD*, Numéro 3, pp 163-187, 2005.
- Ralaivola L., *Introduction aux arbres de décision*. Cours donné à l'université de Provence, 2005.
- Shih Y., Families of Splitting Criteria for Classification Trees, *Statistics and Computing*, 9(4), pp 309-315, 1999.
- Suarez A., Lutsko J., Globally Optimal Fuzzy Decision Trees for Classification and Regression, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(12), pp 1297-1311, 1999.
- Wallace C., Patrick J., Coding Decision Trees, *Machine Learning*, 11, pp 7-22, 1993.
- Wehenkel L., On Uncertainty Measures Used for Decision Tree Induction, in *Proceedings of IPMU*, pp 413-418, 1996.
- Witten I., Frank E., *Data Mining: Practical machine learning tools with Java implementations*, Morgan Kaufmann, San Francisco, 2000.

Zighed D., Rakotomalala R., Graphes d'Induction : Apprentissage et Data Mining, Hermès, 2000.

Chapitre V : Les diagrammes d'influence

1. Introduction

Tout comme les arbres de décision, les diagrammes d'influence (DI) fournissent un moyen graphique pour la modélisation et la résolution des problèmes de décision dans le but de simplifier le processus décisionnel. Ainsi, son rôle se compare à celui que joue le plan pour un architecte. Il permet de raisonner avec rigueur sur le modèle avant même sa mise en chantier. Il permet aussi de s'assurer que l'on n'a oublié aucune variable importante lors de la modélisation.

Un diagramme d'influence est également un outil de communication. Il permet aux différents acteurs de la prise de décision de s'entendre sur un modèle rigoureux. Le diagramme synthétise l'information contenue dans le modèle final.

2. Composants d'un DI

Un diagramme d'influence est un réseau de graphe orienté représentant les croyances et les préférences d'un seul décideur sur une séquence de décisions à prendre dans l'incertitude (Howard, 1984). Il inclue des variables, des formules et des flèches (arcs).

2.1. Les variables

Elles représentent les nœuds du diagramme. On distingue quatre types de variables :

- **De décision** : ce sont les variables sur lesquelles le décideur peut agir de manière directe. Elles sont représentées par des carrés.
- **Résultats** : ce sont des variables directement appréciées par le décideur pour la prise de décision. Les résultats sont représentés par des losanges.
- **Intermédiaires (aléatoires)** : ce sont des variables qui servent de liens entre les variables de décision et celles de résultat à travers des formules. Elles sont représentées par des cercles.
- **Paramètres (nœuds d'utilité)** : Il s'agit de constantes qui entrent dans les calculs. Les constantes peuvent se présenter sous la forme de tableaux à deux ou à plusieurs dimensions. Elles sont représentées par des triangles.

2.2. Les formules

Il s'agit de formules mathématiques ou autres qui mettent en jeu deux ou plusieurs variables. Les formules n'ont pas de représentation graphique sur le diagramme mais sont ajoutées en marge.

2.3. Les flèches (arcs)

Elles traduisent les influences des variables les unes sur les autres. Une flèche partant d'une variable X vers Y signifie que Y est obtenu par une formule à partir de la variable X. Autrement dit, elle signifie que X influence Y.

L'ensemble des arcs se partitionne en deux sous-ensembles : un ensemble d'*arcs fonctionnels* allant d'une variable de décision ou d'une variable intermédiaire vers une variable intermédiaire ou un nœud d'utilité, et un ensemble d'*arcs informationnels* allant d'une variable de décision ou d'une variable intermédiaire vers une variable de décision.

Règles pour utiliser les arcs comme des relations :

Tout comme les nœuds, les arcs aussi jouent un rôle important dans les diagrammes d'influence. La présence d'une flèche indique la signification d'une relation entre le prédécesseur et le successeur.

Un arc peut être impliqué dans deux types de relations, soit dans une pertinence soit dans une séquence. Les différentes interprétations existantes pour les relations sont les suivantes :

- Une flèche pointant vers un nœud intermédiaire (aléatoire) désigne une **relevance**, c'est-à-dire que le nœud prédécesseur est important pour déterminer les chances associées au nœud aléatoire. Dans la figure ci-dessous, la flèche du nœud B au nœud C signifie que l'alternative choisie dans le nœud de décision B a une relevance importante pour établir les chances associées à l'évènement aléatoire. Par exemple, supposons que le nœud C représente la santé publique et B la décision de faire des campagnes de vaccination. Les probabilités qu'une ville ait une bonne santé publique dépendent directement de la décision prise par le gouvernement de faire ou de ne pas faire des campagnes de vaccination. Alors, le choix pris dans la décision B est relevant pour déterminer le résultat de C.
- Un arc pointant vers un nœud de gain désigne lui aussi un arc de relevance. Il indique que le nœud du résultat final dépend du résultat de ses prédécesseurs. Dans le cas de la figure suivante, le résultat final F dépend de la décision D et de l'évènement aléatoire E.
- Une ou plusieurs flèches pointant vers un nœud de décision forment en fait une **séquence**. Quand il existe un arc d'un nœud décision vers un autre nœud décision, cela signifie qu'au moment de la prise de la deuxième décision, la première a été déjà prise. Si un arc part d'un nœud aléatoire vers un nœud décision, cela signifie qu'au moment de la prise de la décision, l'évènement incertain a déjà été résolu. Dans l'exemple de la figure ci-dessous, au moment où le décideur fait son choix I, il connaît déjà le résultat de G, et H n'est plus un évènement aléatoire ou incertain.

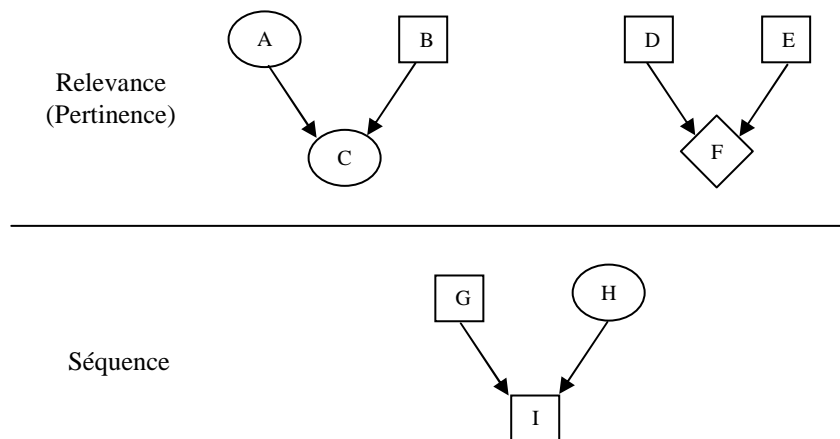


Figure V.1. Interprétations de relations.

Remarque :

Les diagrammes d'influence ne contenant que des variables aléatoires sont cités sous plusieurs noms, tel que les réseaux de croyance.

Exemple : Exemple du parapluie

Cet exemple a été décrit par (Shachter, 1992). Considérons le diagramme d'influence donné par la figure V.2.a qui représente la décision de porter un parapluie au travail. Notre objectif est de maximiser notre satisfaction qui dépend du temps et du fait de porter le parapluie. Notre décision est de porter le parapluie et l'incertitude clé est le temps, que nous ne voyons pas jusqu'à ce que nous faisons notre décision. Nous pouvons consulter les prévisions météo avant de prendre la décision. Les prévisions dépendent du temps, toutefois, elles peuvent ne pas fournir des informations utiles. Un problème plus compliqué est représenté par le diagramme d'influence donné à travers la figure V.2.b. Nous avons déjà observé la météo dans le journal, qui dépend du temps. Maintenant nous cherchons à choisir la station de télévision à regarder pour avoir une idée sur les prévisions météo, qui dépend maintenant à la fois du temps et de la station de télévision. Nous saurons quelle station de télévision, nous avons choisi et ses prévisions avant de décider de porter le parapluie.

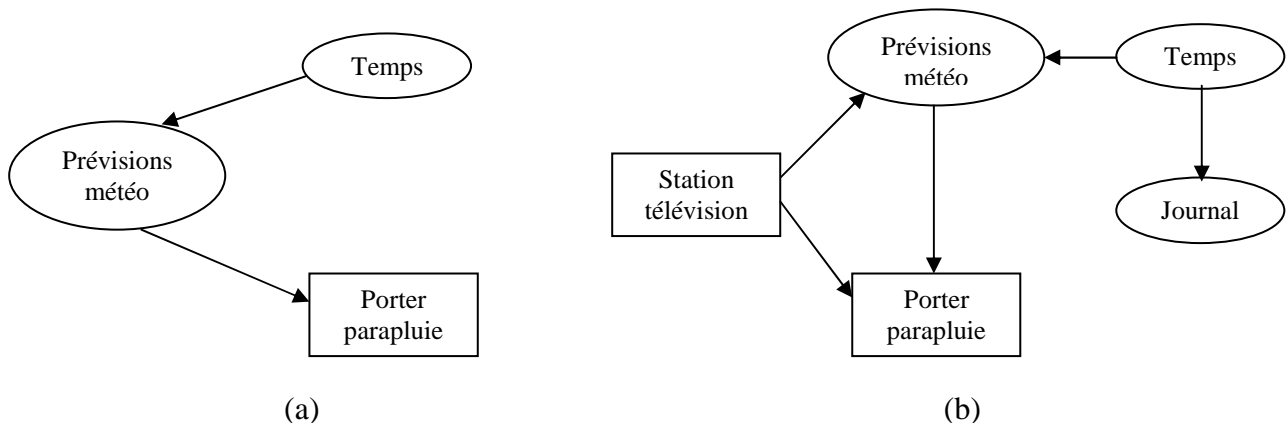


Figure V.2. Exemple du parapluie en diagramme d'influence.

3. Conception d'un diagramme d'influence

Il n'existe pas vraiment de stratégie pour construire des diagrammes d'influence, mais généralement ils sont construits de manière incrémentale à partir des variables résultats ou des variables de décision. Dans ce cas, les étapes de développement sont les suivantes :

Étape 1: la décision à prendre

Il s'agit de décider où commence et où finit le diagramme d'influence. Ainsi, nous devons définir la décision à prendre (il s'agit de la raison d'être du modèle). Il convient également de définir le critère principal d'évaluation.

Étape 2: les différentes variables

Il s'agit maintenant de bâtir le modèle en refaisant le chemin inverse à partir du critère de décision. Il faut représenter toutes les variables intermédiaires qui ont un impact important dans l'évaluation de la décision à prendre.

Étape 3: les liens entre variables

Il faut maintenant définir le type de lien entre les variables représenté par des flèches. À ce stade, nous devons également définir le type de chaque variable intermédiaire. Certaines de ces variables sont dites *exogènes* et ne peuvent être prédites avec certitude (comme les quantités vendues ou le prix des matières premières par exemple). L'entreprise ne peut pas les fixer ni les connaître à l'avance. Il est possible de les anticiper ou encore d'en simuler le comportement aléatoire. Tandis que d'autres sont certaines ou *déterminées avec certitude*.

Étape 4: la relecture

La relecture du diagramme permet de s'assurer de sa cohérence et que le concepteur n'a omis aucune variable essentielle. Il est aussi à éviter de surcharger le diagramme avec des variables redondantes ou peu utiles à la compréhension comme il ne doit pas comporter de cycle.

Remarque :

Un diagramme d'influence doit satisfaire la condition structurelle suivante : « il existe un chemin (comportant des arcs fonctionnels et informationnels) connectant tous les nœuds représentant les variables de décision ».

4. Caractéristiques des diagrammes d'influence

Les diagrammes d'influence sont révélés être un outil efficace pour communiquer non seulement des modèles de décision chez les analystes de décision et les décideurs, mais aussi pour la communication entre l'analyste et l'ordinateur. Ceci provient d'une combinaison des ingrédients suivants (Talman, 1985) :

- ✓ L'objet dominant représenté par le diagramme d'influence est la définition mathématique de la dépendance et de la structure d'information du problème.
- ✓ Cette structure est, en même temps, l'information la plus importante sur le problème à des fins d'analyse et une représentation naturelle et intuitive des aspects structurels du problème que le décideur trouve plus importante.
- ✓ Cette structure est capturée dans le diagramme d'influence sous forme de graphique. Ainsi, les techniques de la théorie des graphes peuvent être utilisés dans l'analyse du modèle de décision. En outre, des représentations graphiques sont naturelle et intuitive pour le décideur.
- ✓ Le diagramme d'influence est d'une hiérarchie à deux niveaux. Cette hiérarchie permet la prise en compte effective des aspects critiques de la structure du modèle dans le niveau supérieur de la hiérarchie, un cadre graphique épurée par les détails quantitatifs. Cependant, ces derniers sont capturés dans le deuxième niveau de la hiérarchie de telle sorte que le diagramme d'influence soit une représentation complète du modèle de décision qui contient toutes les informations nécessaires pour l'analyse.
- ✓ L'analyse des diagrammes d'influence a été automatisé. Cette automatisation est grandement facilitée par le calcul et la structure de données des propriétés symboliques du langage de programmation Lisp.

5. Références bibliographiques

Howard, R. A. et Matheson, J. E., Influence Diagrams. In R. A. Howard and J. E. Matheson (Ed.), The principles and applications of Decision Analysis Menlo Park, CA: Strategic Decisions Group, 1984.

Shachter, R. D. and Peot, M. A., Decision Making Using Probabilistic Inference Methods. In Uncertainty in Artificial Intelligence: Proceedings of the Eighth Conference, pp. 276-283. San Mateo, CA: Morgan Kaufmann, 1992.

Talman, J. A., Decision Process in Influence Diagrams: Formulation and Analysis. Ph.D. Thesis, Engineering-Economic Systems Dept., Stanford University, 1985.

Voir aussi :

Gonzales. C. Diagrammes d'influence et CP-nets, Transparents de cours, Module MGDE, Master IAD, LIP6 – Paris 6 Université, France, 2012-2013.

Howard. R. and Matheson. J. Influence Diagrams. In Readings on the Principles and Applications of Decision Analysis, pp 721–762. Strategic Decisions Group, Menlo Park, CA, USA, 1984.

Jeantet. G. and Spanjaard. O. Optimisation de l'utilité espérée dépendant du rang dans les diagrammes d'influence. 10ème Congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF 2009), 2009.

Magnan. J-Ch. and Wuillemin. P-H. Improving Decision Diagrams for Decision Theoretic planning. Florida Artificial Intelligence Research Society Conference, pp 621-626, 2013.

Sedki. K. and Delcroix. V. Diagrammes d'influence pour l'aide à la décision multi-critère. Actes du 12ème congrès annuel de la Société française de Recherche Opérationnelle et d'Aide à la Décision, janvier 2011.

Shachter. R. Evaluating Influence Diagrams. Operations Research, 34(6): pp 871–882, 1986.

Shachter, R.D., Probabilistic inference and influence diagrams. In Operating Research, volume 36, pages 589–604, 1988.

Chapitre VI : Les réseaux bayésiens

1. Introduction

L'information n'est pas la connaissance. Plus la technologie facilite le stockage et l'échange d'informations, plus se pose la question de l'analyse et de la synthèse de ces informations. Deux types d'approches connaissent donc un intérêt croissant : la statistique qui autorise le passage de l'observation à la loi, et l'intelligence artificielle, qui permet aux ordinateurs de traiter la connaissance plutôt que l'information.

Les réseaux bayésiens sont le résultat du mariage de ces deux approches et sont aujourd'hui l'un des formalismes les plus prometteurs pour l'acquisition, la représentation et l'utilisation de connaissances par des ordinateurs.

Ceux sont des système représentant la connaissance et permettant de calculer des probabilités conditionnelles apportant des solutions à différentes sortes de problématiques. La structure de ce type de réseaux est simple : un graphe dans lequel les nœuds représentent des variables aléatoires, et les arcs (le graphe est donc orienté) reliant ces dernières sont rattachées à des probabilités conditionnelles.

Notons que le graphe est acyclique : il ne contient pas de boucle. Les arcs représentent des relations entre variables qui sont soit déterministes, soit probabilistes. Ainsi, l'observation d'une ou plusieurs causes n'entraîne pas systématiquement l'effet ou les effets qui en dépendent, mais modifie seulement la probabilité de les observer. L'intérêt particulier des réseaux bayésiens est de tenir compte simultanément de connaissances a priori d'experts (dans le graphe) et de l'expérience contenue dans les données (Parent, 2006).

Donc, Un réseau bayésien est défini par (Leray, 2008) :

- ✓ la description qualitative des dépendances (ou des indépendances conditionnelles) entre des (variables graphe orienté (DAG)).
- ✓ la description quantitative de ces dépendances (probabilités conditionnelles (CPD)).

Ils sont appliqués aujourd'hui dans les domaines de la santé (diagnostic, localisation de gènes), de l'industrie (contrôle de véhicule autonomes), de la défense (fusion de données), dans le marketing (data-mining), dans l'informatique et les réseaux (agents intelligents) et dans la gestion de la connaissance. Pour toutes ces applications, les réseaux bayésiens ont la capacité de découvrir des relations, de les formaliser, de les exploiter et de les faire évoluer dans des environnements de décision où l'incertitude est la règle.

2. Probabilités

Le domaine des réseaux bayésiens a comme particularité d'allier deux champs différents des mathématiques dans le but de représenter l'incertitude : la théorie des graphes, d'une part, qui fournit le cadre nécessaire pour une modélisation qualitative des connaissances ; et la théorie des probabilités, d'autre part, qui permet d'introduire une information quantitative dans ces connaissances.

2.1. Définitions principales

Définition 1 : (Probabilité)

Considérons une expérience dont l'ensemble fondamental est Ω . On définit sur l'ensemble des évènements une fonction qui à chaque évènement associe un poids, appelé *probabilité*, compris entre 0 et 1 : plus le poids est proche de 0 moins l'évènement risque de se réaliser, plus le poids est proche de 1 plus l'évènement se réalisera souvent (Feller, 1968).

Formellement, une probabilité P sur Ω est une application sur l'ensemble des évènements telle que :

- 1) $P(\Omega) = 1$
- 2) pour tout évènement A , $0 \leq P(A) \leq 1$
- 3) pour toute suite A_1, A_2, \dots d'évènements disjoints (i.e, pour tout $i \neq j, A_i \cap A_j = \emptyset$), on a :

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Une probabilité est une mesure dans le sens où plus un évènement est grand, plus sa probabilité est importante.

Il est à noter qu'un évènement sur Ω , appelé l'*univers*, est une sous-partie de Ω . Toute expérience aléatoire comprend un évènement certain et un évènement impossible. L'évènement impossible noté \emptyset , est le sous-ensemble qui ne contient pas d'éléments. L'évènement certain est le sous-ensemble qui contient tous les éléments, c'est l'ensemble fondamental Ω lui même.

Exemple 1 :

On jette un dé équilibré à 6 faces. On regarde le nombre qui apparaît sur la face supérieure du dé. Les issues possibles sont : 1, 2, 3, 4, 5 et 6.

L'évènement « obtenir le chiffre 7 » est un évènement impossible.

L'évènement « obtenir le chiffre 1, 2, 3, 4, 5 ou 6 » est un évènement certain.

▪ Définition 2 : (Variable aléatoire)

Une variable aléatoire est une fonction X définie sur Ω : $X : \begin{cases} \Omega \rightarrow D_X \\ \omega \mapsto X(\omega) \end{cases}$

Pour $x \in D_X$, on note alors $\{X = x\}$ l'évènement $\{\omega \in \Omega \mid X(\omega) = x\}$

D_X est le domaine de définition de X .

Exemple 2 :

Pour étudier la distribution de probabilité de la somme du tirage de deux dés, il suffit de définir une variable aléatoire représentant cette somme, ce qui permet de manipuler beaucoup plus facilement les évènements correspondants.

| D_X | ...1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13... |
|----------------|-------------|----------------|----------------|-------------------------|----------------------------------|---|--|---|----------------------------------|-------------------------|----------------|----------------|-------------|
| $\{X = x\}$ | \emptyset | (1,1) | (1,2) (2,1) | (1,3) (2,2) (3,1) | (1,4) (2,3) (3,2) (4,1) | (1,5) (2,4) (3,3) (4,2) (5,1) | (1,6) (2,5) (3,4) (4,3) (5,2) (6,1) | (2,6) (3,5) (4,4) (5,3) (6,2) | (3,6) (4,5) (5,4) (6,3) | (4,6) (5,5) (6,4) | (5,6) (6,5) | (6,6) | \emptyset |
| $P(\{X = x\})$ | 0 | $\frac{1}{36}$ | $\frac{1}{18}$ | $\frac{1}{12}$ | $\frac{1}{9}$ | $\frac{5}{36}$ | $\frac{1}{6}$ | $\frac{5}{36}$ | $\frac{1}{9}$ | $\frac{1}{12}$ | $\frac{1}{18}$ | $\frac{1}{36}$ | 0 |

Table VI.1. Distribution des évènements élémentaires en fonction d'une valeur aléatoire. (Naïm, 2004)

2.2. Probabilités sur plusieurs variables

Une variable aléatoire est donc un moyen pour condenser une information pertinente sur un univers. Toutefois, les problèmes traités dans la réalité nécessitent souvent pour leurs descriptions plus d'une variable. Pour reprendre l'exemple du tirage de deux dés, la somme des deux tirages est une information intéressante, mais la valeur de chacun des deux tirages est une autre information qui peut s'avérer nécessaire. L'étape suivante est bien sûr d'avoir le moyen de croiser ces différentes sources d'informations.

■ Probabilités jointes

En pratique, on décrit souvent un système grâce à un ensemble de paramètres qui permet de le caractériser à tout moment. Suivant la disponibilité ou non des informations sur les valeurs de ces paramètres, on distingue, respectivement, deux types de systèmes : un système déterministe et un système probabiliste. Dans ce dernier cas et dans le but de décrire un tel système, il faut tenter de lui adjoindre une probabilité sur les différentes variables correspondantes aux différents paramètres.

Définition 3 : (Probabilités jointes)

Soient A et B deux variables aléatoires sur le même univers Ω . On parle alors de probabilité pour la fonction définie sur $D_A \times D_B$ par :

$$P_{AB} : \begin{cases} D_A \times D_B \rightarrow [0,1] \\ (a,b) \mapsto P_{AB}(a,b) = P(\{A=a\} \cap \{B=b\}) = P(\{\omega \in \Omega \mid A(\omega)=a \wedge B(\omega)=b\}) \end{cases} \quad \dots(1)$$

Cette définition peut être étendue à tout ensemble $U = \{X_1, \dots, X_n\}$ de variables aléatoires définies sur le même univers Ω .

$$P_U : \begin{cases} \bigotimes_{i \in \{1, \dots, n\}} D_{X_i} \rightarrow [0,1] \\ u = (x_1, \dots, x_n) \mapsto P_U(u) = P\left(\bigcap_{i \in \{1, \dots, n\}} \{X_i = x_i\}\right) = P\left(\{\omega \in \Omega \mid \bigwedge_{i \in \{1, \dots, n\}} X_i(\omega) = x_i\}\right) \end{cases} \quad \dots(2)$$

■ Probabilités marginales

Réciproquement, la *probabilité marginale* permet de retrouver la probabilité jointe de chacun des sous-ensembles de variables d'une probabilité jointe.

Propriété 4 : (Marginalisation)

Soient :

- U un ensemble fini, non vide de variables aléatoires,
- $V \subset U$ non vide,
- $V' = U \setminus V$ et $P(U)$ la probabilité jointe sur les variables de U.

On appelle *marginalisation* de P sur V la fonction :

$$\forall v \in D_V, P(v) = \sum_{v' \in D_{V'}} P(v, v') \quad \dots(3)$$

Cette fonction correspond à la probabilité jointe des variables de V.

L'opération de marginalisation peut être généralisée à toute fonction f sur un ensemble de variable U. La notation usuelle pour cette opération est $[f]^{\downarrow V}$ où $V \subset U$. Donc, la propriété (3) peut s'écrire fonctionnellement :

$$\forall V \subset U, P(V) = [P(U)]^{\downarrow V} = \sum_{v' \in U \setminus V} P(V, v') \quad \dots(4)$$

Exemple 3 :

Soit, par exemple, deux variables aléatoires T et L, dont la probabilité jointe suit le tableau suivant :

| P(l,t) | l_1 | l_2 |
|--------|--------|--------|
| t_1 | 0.0578 | 0.0782 |
| t_2 | 0.1604 | 0.0576 |
| t_3 | 0.5118 | 0.1342 |

Par marginalisation, on peut obtenir :

$$P(L = l_1) = P(L = l_1, T = t_1) + P(L = l_1, T = t_2) + P(L = l_1, T = t_3) = 0.73$$

▪ **Probabilités conditionnelles**

La notion de *probabilité conditionnelle* est primordiale en calcul des probabilités. En effet, soient :

- ✓ ω un évènement de l'univers Ω ,
- ✓ $P(\omega)$ sa probabilité associée à des conditions de réalisation.

Dans l'exemple 2, on ne peut étudier la distribution de probabilité de la somme du tirage de deux dès sauf au cas où on suppose que les deux dès ont été tirés. Dans ce cas, les évènements « ne tirer aucune dès » et « tirer un seul dès » sont les évènements impossibles.

Ainsi, toute probabilité doit impliquer la prise en compte d'un certain contexte de réalisation, sans lequel elle n'existe pas. Autrement dit, toute probabilité est conditionnelle à un certain contexte. La question « quelle est la probabilité de A » devrait toujours être comprise comme « Etant donné le contexte ϵ , quelle est la probabilité de A ? », ce qui se note $P(A | \epsilon)$.

Formellement :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Théorème de Bayes :

Si (Ω, Z, P) est un espace probabilisé, et $(A_i)_{i \in I}$ un ensemble complet d'évènements de Z, c'est-à-dire tel que :

$$1) \forall (i, j), i \neq j, A_i \cap A_j = \emptyset$$

$$2) \bigcup_{i \in I} A_i = \Omega$$

$$\text{Alors : } \forall B \in Z, P(B) = \sum_{i \in I} P(B | A_i) \cdot P(A_i)$$

De ce théorème et de la définition de probabilité conditionnelle, nous déduisons :

$$P(A_k | B) = \frac{P(B | A_k) \cdot P(A_k)}{\sum_{i \in I} P(B | A_i) \cdot P(A_i)} \quad \dots \text{ Bayes-3}$$

$$\text{et } \forall B, C \in Z, P(B | C) = \sum_{i \in I} P(B | A_i \cap C) \cdot P(A_i | C)$$

▪ **Définition : Indépendance**

Soit : (Ω, Z, P) un espace probabilisé,
soient : A et B deux éléments de Z.

On dit que A et B sont *indépendants* si et seulement si :

$$P(A \cap B) = P(A) \cdot P(B)$$

▪ **Définition : Indépendance conditionnelle**

Soit : (Ω, Z, P) un espace probabilisé,

soient : A, B et C trois éléments de Z.

On dit que A et B sont *indépendants conditionnellement* à C si et seulement si :

$$P(A \cap B | C) = P(A | C).P(B | C)$$

▪ **Définition (Loi fondamentale)**

$$P(a,b) = P(a | b).P(b)$$

▪ **Théorème**

Si P(b) est positive, alors :

$$P(a | b) = \frac{P(b | a).P(a)}{P(b)} \quad \dots \text{ Bayes-1}$$

Plus généralement :

$$P(a | b,c) = \frac{P(b | a,c).P(a | c)}{P(b | c)} \quad \dots \text{ Bayes-2}$$

3. Les réseaux bayésiens

Un réseau bayésien est défini par :

- un graphe acyclique orienté G, $G = (V, E)$, où V est l'ensemble des nœuds de G, et E l'ensemble des arcs de G;
- un espace probabiliste fini (Ω, Z, p) ;
- un ensemble de variables aléatoires correspondant aux nœuds du graphe définies sur (Ω, Z, p) , tel que :

$$p(V_1, V_2, \dots, V_n) = \prod_{i=1, n} p(V_i | C(V_i))$$

où $C(V_i)$ est l'ensemble des causes (parents) de V_i dans le graphe G.

Donc et dans le but de résoudre un problème donné, un réseaux bayésien présente à la fois un raisonnement qualitatif et quantitatif. Le premier est assuré par l'ensemble de contraintes de traitement de la circulation de l'information sur la représentation graphique du problème. Le graphe impliqué est un graphe de causalité. Le raisonnement quantitatif permet la validation du raisonnement qualitatif en exploitant les notion de probabilités.

4. Graphe de causalité (Représentation qualitative)

4.1. Représentation de la causalité

La représentation graphique ci-dessous peut être intuitivement traduite en : A a une influence sur B. Ainsi et du point de vue du sens commun, A sera la cause de l'effet B.



En effet, la connaissance sur A détermine la connaissance sur B, ou autrement dit, sachant avec certitude l'état de A, je peut en déduire l'état de B avec certitude. Ceci est équivalent au fait de dire que l'information circule de A vers B en suivant le sens d'orientation de la flèche.

Mais en réalité, malgré que la flèche soit orientée de A vers B, elle peut fonctionner dans les deux sens, et ce même si la relation causale est stricte.

Exemple :

Nous allons reprendre l'exemple de (Naïm, 2004) dans le but de prouver que sur une flèche l'information peut circuler dans les deux sens.

Etant donnée que la relation causale soit l'implication logique $A \Rightarrow B$, cette relation signifie que si A est vrai, B l'est également. Si A est faux, B peut être vrai ou faux. Ceci est bien explicite sur la table ci-dessous énumérant les configurations possibles de A et B suivant la relation causale considérée ($A \Rightarrow B$). Cela veut dire que l'information a circulé de A vers B suivant le sens de la flèche.

| A | B |
|---|---|
| V | V |
| F | V |
| F | F |

De plus, cette table nous permet d'affirmer que, si B est faux, A l'est également. Il s'agit simplement de la contraposée logique de $A \Rightarrow B$. Dans ce cas ceci est équivalent au fait que l'information a circulé au sens inverse de la flèche de B vers A.

Donc, nous avons démontré à travers cet exemple que bien que la relation causale $A \Rightarrow B$ soit orientée, elle est réversible de l'effet B vers la cause A, même si elle ne l'est que partiellement.

Conclusion :

« S'il existe une relation causale de A vers B, toute information sur A peut modifier la connaissance que j'ai de B, et, réciproquement, toute information sur B peut modifier la connaissance que j'ai de A ».

En présence d'un graphe plus complexe, il est essentiel de conserver à l'esprit que l'information ne circule pas seulement dans le sens des flèches.

4.2. Circulation de l'information dans un graphe de causalité

Dans cette section, nous considérons toujours qu'un graphe de causalité est une simple liaison des causes aux effets.

4.2.1. Exemple

Considérons l'exemple, extrêmement classique dans la littérature proposé par Pearl dans (Pearl, 1988) sur les réseaux bayésiens :

« Ce matin-là, alors que le temps est clair et sec, M. Holmes sort de sa maison. Il s'aperçoit que la pelouse de son jardin est humide. Il se demande alors s'il a plu pendant la nuit, ou s'il a simplement oublié de débrancher son arroseur automatique. Il jette alors un coup d'œil à la pelouse de son voisin, et s'aperçoit qu'elle est également humide. Il en déduit alors qu'il a probablement plu, et il décide de partir au travail sans vérifier sans arroseur automatique ».

D'après cet exemple, l'ensemble d'informations utiles regroupe les quatre informations suivantes : Arroseur (A), Pluie (P), Jardin humide (J), Jardin du voisin humide (W). Ainsi, le graphe de causalité modélisant ce contexte sera composé de 4 nœuds. C'est celui illustré à travers la figure VI.1.

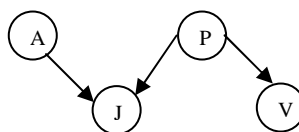


Figure VI.1. Graphe de causalités modélisant l'exemple de l'arroseur de Pearl.

Le graphe précédent énumère trois liaisons qui sont :

- La liaison (P - J) indiquant que l'herbe du jardin de M. Holmes est humide du fait qu'il a plu pendant la nuit.
- La liaison (P - V) indiquant que l'herbe du jardin du voisin de M. Holmes est humide du fait qu'il a plu pendant la nuit.
- La liaison (A - J) indiquant que l'herbe du jardin de M. Holmes est humide du fait qu'il a oublié de débrancher son arroseur automatique.

Ce graphe peut être utilisé de la manière suivante par M. Holmes pour arriver à conclure sur l'état de son arroseur :

1) La connaissance de J renforce la croyance en l'une des deux causes A ou P, c-à-d, les deux causes sont *à priori* également plausibles.

2) La connaissance de V augmente la croyance en la cause P. Par rapport à l'état (1) précédent, la cause A devient moins plausible.

D'après le raisonnement explicité en (1) et (2), on peut déduire que l'arroseur automatique était à l'arrêt du fait que la pelouse du jardin du voisin de M. Holmes était également humide. Ceci est équivalent à une circulation de l'information de V vers A.

Conclusion :

Sur un graphe de causalité, l'information peut circuler en suivant des chemins qui peuvent paraître à première vue contre-intuitive.

4.2.2. Indépendance conditionnelle et d-séparation

Les connexions entre les nœuds définissent des lois de circulation de l'information dans le graphe. On distingue trois types de connexions :

- *connexion en série* (Figure VI.2.a) : l'information ne peut circuler entre X et Y que si la valeur de Z n'est pas connue, sinon c'est directement la connaissance sur le nœud Z qui influe ;
- *connexion divergente* (Figure VI.2.b) : comme précédemment, l'information ne peut circuler entre X et Y que si la valeur de Z n'est pas connue ;
- *connexion convergente ou connexion en V* (Figure VI.2.c) : l'information ne peut circuler entre X et Y que si la valeur de Z est connue.

Ainsi, la circulation de l'information à l'intérieur d'un graphe dépend du type des connexions, plutôt que du sens des flèches.

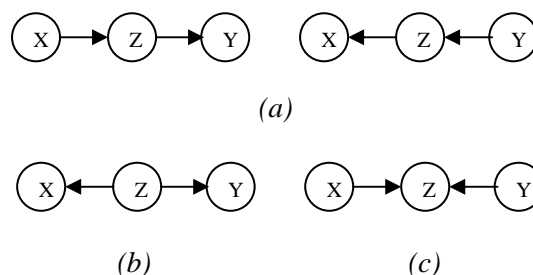


Figure VI.2. Types de connexions entre trois nœuds, (a) en série, (b) divergente, (c) convergente.

Dans (Naïm, 2004), les auteurs ont essayé de déterminer si l'information peut circuler de X à Y, en considérant chaque fois un petit exemple (Voir table VI.2).

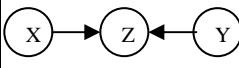
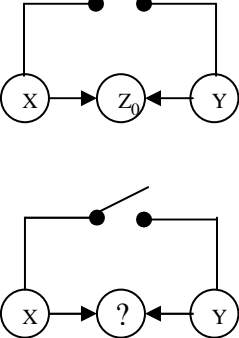
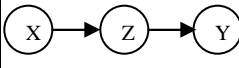
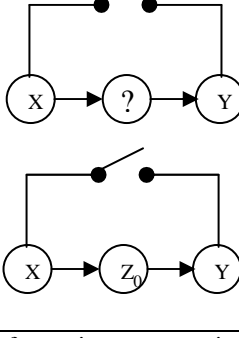
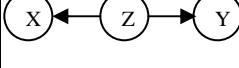
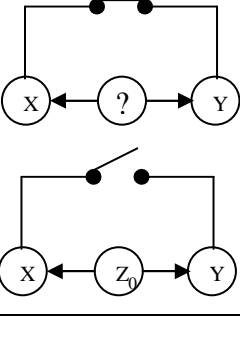
| Graphe | Propriété | Exemple |
|---|--|--|
|  | <p>L'information ne peut circuler de X à Y que si Z est connu</p>  | <p>X = tremblement de terre Y = cambriolage Z = alarme</p> <p>Le fait qu'il y ait eu un tremblement de terre dans le voisinage (X) n'a aucun lien <i>à priori</i> avec le fait que ma maison ait été cambriolée (Y). En revanche, si mon alarme s'est déclenchée (Z), j'ai tendance à croire que je viens d'être cambriolé (Y). Si maintenant j'apprends qu'il vient d'y avoir un tremblement de terre (X) dans le voisinage, je suis rassuré sur l'éventualité d'un cambriolage (Y).</p> |
|  | <p>L'information ne peut circuler de X à Y que si Z n'est pas connu.</p>  | <p>X = ensoleillement Y = prix du blé Z = récolte</p> <p>Si la saison a été ensoleillée (X), la récolte sera abondante (Z). Si la récolte est abondante, le prix du blé est bas (Y). Si je sais déjà que la récolte a été abondante (Z), le fait de connaître l'ensoleillement (X) ne m'apprend plus rien sur le prix du blé (Y).</p> |
|  | <p>L'information ne peut circuler de X à Y que si Z n'est pas connu.</p>  | <p>X = la pelouse de mon jardin est humide Y = la pelouse de mon voisin est humide Z = il a plu cette nuit</p> <p>Si la pelouse de mon jardin est humide (X), j'ai tendance à croire qu'il a plu cette nuit (Z), et donc que la pelouse de mon voisin sera aussi humide (Y). Si en revanche je sais qu'il a plu cette nuit (Z), je peux affirmer que la pelouse du jardin de mon voisin sera humide (Y), et l'information que je peux avoir sur l'état de ma propre pelouse (X) n'y change rien.</p> |

Table VI.2. Exemples illustratifs de la circulation de l'information suivant les types de connexions de nœuds. (Naïm, 2004)

La notion de *d-séparation* est essentielle pour le calcul des probabilités a posteriori car elle permet de définir l'indépendance conditionnelle entre certains nœuds. Soient X, Y et Z trois nœuds du graphe, on dit que X et Y sont d-séparés par Z (et on note $\langle X | Z | Y \rangle$) si, pour tous les chemins entre X et Y, l'une au moins des deux conditions suivantes est vérifiée :

- ✓ Le chemin converge en un nœud W, tel que $W \neq Z$, et Z n'est pas un descendant de W.
- ✓ Le chemin passe par Z avec une connexion divergente ou en série.

Dans la figure VI.2 (a) et (b), X et Y sont d-séparés par Z, alors que dans la figure VI.2.(c) Z empêche la d-séparation. En d'autres termes, lorsque Z est connu (on dit observé), l'information circule entre X et Y dans les deux sens. La définition de d-séparation s'étend à des sous-ensemble de variables.

Le théorème fondamental des réseaux bayésiens est défini comme suit : « si X et Y sont d-séparés par Z, alors X et Y sont indépendants sachant Z » :

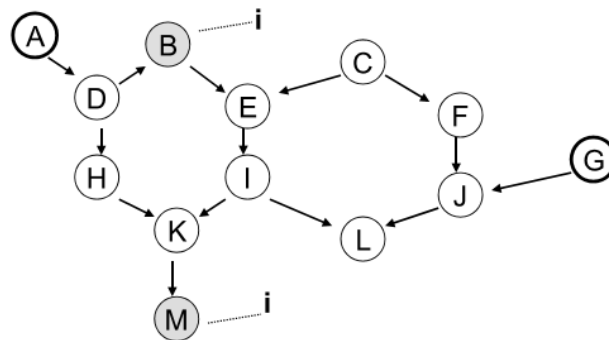
$$\langle X | Z | Y \rangle \Rightarrow P(X|Y, Z) = P(X|Z)$$

Ainsi la structure du graphe d'un réseau bayésien encode un certain nombre d'indépendances conditionnelles. Cette propriété est exploitée dans les algorithmes d'inférence pour réduire la partie du graphe à considérer.

En revanche, la réciproque est fautive (Ben Mrad, 2015) : certaines indépendances conditionnelles, présentées dans la distribution de probabilité, peuvent ne pas se traduire dans le graphe. Certains graphes sont "meilleurs" que d'autres pour une même distribution de probabilité jointe sur un ensemble de variables X, car ils représentent plus d'indépendances conditionnelles. Par ailleurs, différents graphes peuvent encoder les mêmes indépendances conditionnelles, puisque seules les V-structures se différencient des deux autres types de connexion.

Exemple :

Soit le graphe de causalité suivant :



Sachons que B et M soient instanciés (connus), donc, A et G sont d-séparés.

5. Représentation probabiliste associée (quantitative)

Le but de cette section est de présenter une formalisation à base de probabilités de la représentation qualitative illustrée dans la section précédente. Précisément, on cherche à valider les principes qualitatifs de dépendances et d'indépendances et ceux de circuits d'information et d'indépendances. Ces deux cas seront illustrés à travers des études de cas formalisant les mêmes notions précédemment vues sur les exemples ultérieurs (exemple de l'implication logique et celui du jardin de Holmes). Toutefois il faudra tout d'abord présenter la manière de transposer les informations utiles présentées sur le graphe en variables exploitables via des lois de probabilités. C'est l'objectif de la première sous section.

Une fois les résultats probabilistes seront obtenus, nous présenteront l'équivalence entre la représentation qualitative et la représentation quantitative.

5.1. Transposition qualitative / quantitative

Ici nous présentons les règles assurant la transposition ou la formalisation de la représentation graphique en une représentation probabiliste.

5.1.1. Définition des variables

Règle 1 (Espace probabilisé, évènement) :

Soient A et B deux nœuds dans un graphe causal pouvant prendre chacun la valeur « vrai » ou « faux ». L'espace probabilisé, noté E , correspondant sera constitué des couples suivants :

$$E = \{(A = V, B = V), (A = V, B = F), (A = F, B = V), (A = F, B = F)\}$$

Où Chaque couple est appelé un évènement.

Règle 2 (Variable aléatoire) :

Considérons le même contexte impliqué dans la formulation de la règle 1, les deux variables A et B seront des *variables aléatoires* sur E définies comme suit :

$$\begin{array}{ll} A((A = V, B = V)) = 1 & B((B = V, A = V)) = 1 \\ A((A = V, B = F)) = 1 & B((B = V, A = F)) = 1 \\ A((A = F, B = V)) = 0 & B((B = F, A = V)) = 0 \\ A((A = F, B = F)) = 0 & B((B = F, A = F)) = 0 \end{array}$$

Où : l'ensemble $\{(A = V, B = V), (A = V, B = F)\}$ est l'image réciproque de 1 par l'application A . Cet ensemble peut être noté simplement par : $A = V$.

Règle 3 (Variables indépendantes) :

Soient A et B deux variables indépendantes. Ainsi, nous aurons :

$$p(A|B) = p(A)$$

En considérant la même définition de variables en règle 2, il est clair que A et B sont indépendantes. On aura donc :

$$\begin{cases} p(A = V / B = V) = p(A = V) \\ p(A = V / B = F) = p(A = V) \\ p(A = F / B = V) = p(A = F) \\ p(A = F / B = F) = p(A = F) \end{cases}$$

5.1.2. Définition des probabilités

Règle 1 :

Soit A une variable n'ayant aucune cause directe.

Le calcul de $p(A)$ revient à calculer $p(A = V)$ et $p(A = F)$.

Règle 2 :

Soit B une variable ayant une seule cause directe A .

Le calcul de $p(B|A)$ revient à calculer $p(A=V, B=V)$, $p(A=V, B=F)$, $p(A=F, B=V)$, $p(A=F, B=F)$.

Règle 3 :

Soit C une variable ayant deux causes directes A et B .

Le calcul de $p(C|A,B)$ revient à calculer $p(C = V/A = V, B = V)$, $p(C = V/A = V, B = F)$, $p(C = V/A = F, B = V)$, $p(C = V/A = F, B = F)$, $p(C = F/A = V, B = V)$, $p(C = F/A = V, B = F)$, $p(C = F/A = F, B = V)$, $p(C = F/A = F, B = F)$.

5.2. Etude de cas N° 01

A travers cette étude de cas nous cherchons à valider la formalisation probabiliste. Pour se faire, nous considérons l'exemple précédemment traité qui celui de l'implication logique $A \Rightarrow B$.

5.2.1. Formalisation qualitative / quantitative

Pour faciliter de raisonner qualitativement, nous allons donner de la signification aux événements A et B .

Nous allons reprendre les mêmes considérations proposées par (Naïm, 2004) en supposant que A et B représentent des événements dans le monde de la finance :

- ✓ A est l'évènement : « L'annonce des chiffres du commerce extérieur américain est supérieur aux attentes du marché ».
- ✓ B est l'évènement : « Le cours du dollar contre l'euro monte ».

De manière qualitative, la signification de l'implication est assurée. En effet si l'annonce des chiffres du commerce extérieur américain est supérieure aux attentes du marché, le cours du dollar contre l'euro va monter par rapport au cours de la veille. Dans le cas contraire, le cours du dollar va être influencé par d'autres causes, et on ne pourra donc rien dire sur son évolution.

Supposons, en premier lieu, que le dollar a enregistré une baisse significative et en n'ayant aucune idée sur les chiffres du commerce extérieur américain, que peut-on déduire sur ces derniers ?

Formalisation probabiliste (quantitative)

Considérons les deux variables A et B représentatives des événements A et B .

▪ Probabilités à priori

Comme nous n'avons aucune information qui confirme A ou \bar{A} , donc les deux cas sont équiprobables, donc on attribue à priori les probabilités résumées dans la table ci-dessous.

| Evènement | Probabilité |
|-----------|-------------|
| $A = V$ | 1/2 |
| $A = F$ | 1/2 |

▪ Probabilités conditionnelles

On admettant que $A \Rightarrow B$ est vraie, on aura :

- ✓ *Cas 1* : Si A s'est réalisé ($A = V$), B se réalisera certainement ($B = V$).
- ✓ *Cas 2* : Si A ne s'est pas réalisé, on pourra rien déduire sur l'état de B ($B = V$ ou $B = F$). Donc les deux cas sont équiprobables.

Les probabilités formalisant ces deux cas sont de type conditionnelles. Elles se résument comme suit:

| | | |
|---------------|-----|-------|
| $B = V A = V$ | 1 | Cas 1 |
| $B = F A = V$ | 0 | |
| $B = V A = F$ | 1/2 | Cas 2 |
| $B = F A = F$ | 1/2 | |

La question posée concernant la valeur des chiffres du commerce extérieur américain sachons que le dollar a enregistré une baisse significative se traduit en une probabilité conditionnelle comme suit: $p(A = V | B = F)$

▪ Calcul

Etant donné que :

$$(A = V) \cap (A = F) = \emptyset$$

$$(A = V) \cup (A = F) = E$$

Donc, les conditions d'application du théorème de Bayes sont vérifiées et on aura :

$$p(A = V / B = F) = \frac{p(B = F / A = V) \cdot p(A = V)}{p(B = F / A = V) \cdot p(A = V) + p(B = F / A = F) \cdot p(A = F)}$$

Donc : $p(A = V / B = F) = 0$

▪ Interprétation

La traduction qualitative du résultat obtenu quantitativement est que le chiffre du commerce extérieur américain a certainement était inférieur aux attentes du marché.

Ce résultat est bien conforme à la discussion de l'exemple de l'implication logique vu dans la section 4.1. En effet, quand B est "faux", A est certainement "faux".

Donc, ce résultat permet de valider la transposition de notre relation causale en termes de probabilités.

Considérons, maintenant, le contraire de la première supposition où le dollar enregistre une forte hausse et on cherche aussi de conclure sur les chiffres du commerce extérieur américain.

Formellement, nous démarrons maintenant avec $B = V$ et nous chercherons à calculer $p(A = V / B = V)$. Vu que les conditions d'application du théorème de Bayes soient toujours vérifiées, nous pouvons écrire :

$$p(A = V / B = V) = \frac{p(B = V / A = V) \cdot p(A = V)}{p(B = V / A = V) \cdot p(A = V) + p(B = V / A = F) \cdot p(A = F)}$$

Ce résultat démontre la véracité de l'implication logique $((A = V) \Rightarrow (B = V))$.

5.2.2. Discussion

Cette section présente un récapitulatif analytique des traitements effectués juste en haut dans la première étude de cas en termes de passage de la formalisation qualitative en formalisation quantitative et vice versa. Nous les résumons en la succession suivante d'opérations.

1) Définition d'un espace probabilisé : les évènements et leurs probabilités sont, ici, définis.

Dans l'étude de cas N° 01, la probabilité $p(A = V) = p(A = F) = 1/2$ est la plus discutable puisqu'elle n'est fondée sur rien d'objectif mais en absence de raisons à priori, aucune croyance n'a été attribuée.

2) Traduction des connaissances certaines : Dans cet étude de cas, et comme nous avons admis que la relation entre A et B était une relation causale stricte, c'est à dire $A \Rightarrow B$, donc l'ensemble des connaissances qui en résulte est celui résumé par la table suivante :

| | A = V | A = F |
|-------|------------|----------|
| B = V | nécessaire | possible |
| B = F | impossible | possible |

Les informations contenues dans cette table peuvent être traduites en probabilités conditionnelles. Les résultats obtenus sont ceux illustrés dans la table ci-dessous. De même que pour le cas critique de l'étape 1, là encore et faute d'absence d'information, la probabilité 1/2 a été attribuée à deux évènements complémentaires.

| | A = V | A = F |
|-------|-------|-------|
| B = V | 1 | 1/2 |
| B = F | 0 | 1/2 |

3) **Formalisation probabiliste du résultat recherché** : Dans cet exemple le résultat recherché exprimé en termes de probabilités est le suivant : $p(A = V / B = F)$ pour la première supposition (respectivement, $p(A = V / B = V)$, pour la deuxième supposition).

4) **Calcul** : Dès que la formalisation soit effectuée et que les conditions d'application du théorème de Bayes soient vérifiés, les probabilités recherchées peuvent être calculées. le théorème de Bayes nous donne immédiatement les probabilités recherchées.

5) **Interprétation du résultat** : Le résultat le plus important de cette formalisation est celui de démontrer l'équivalence entre la représentation graphique et la représentation probabiliste.

Ainsi et une fois que le résultat soit obtenu, nous revenons dans le domaine qualitatif où nous cherchons à interpréter le résultat en termes de conviction : les chiffres du commerce extérieur ont sans doute été bons.

La figure ci-dessous résume l'équivalence entre la représentation qualitative et celle quantitative.

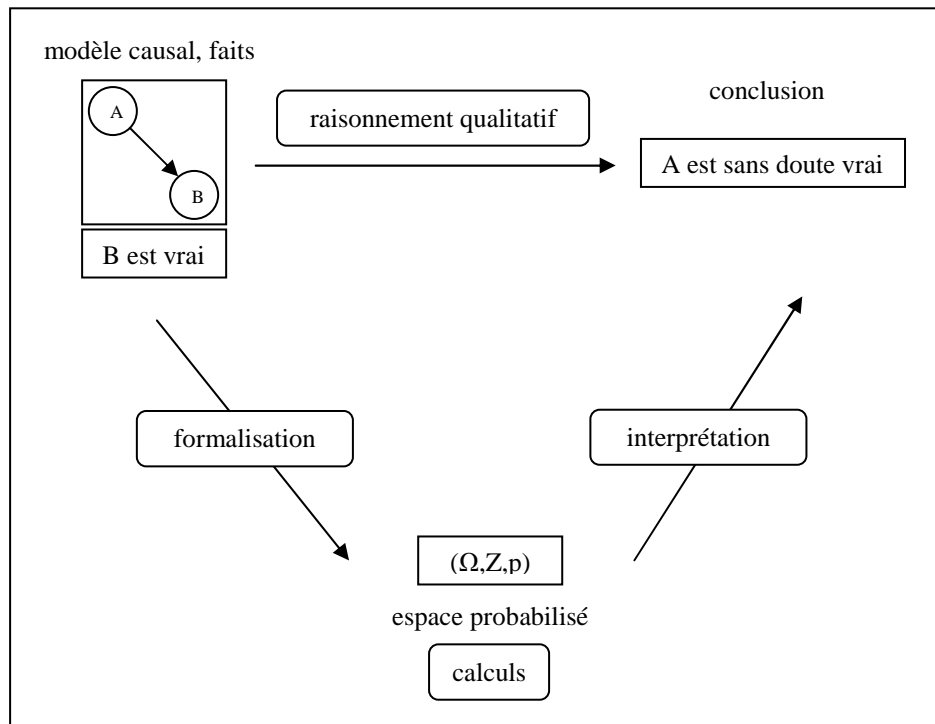


Figure VI.3. Equivalence entre représentation qualitative et quantitative. (Naïm, 2004)

Pour cet exemple, nous arrivons à conclure que l'interprétation des résultats obtenus de manière quantitatives sont équivalents à ceux induits d'un raisonnement qualitatif.

C'est le cas d'ailleurs pour tout problème, malgré que cette équivalence ne peut être prouvée.

5.3. Etude de cas N° 02

Le but recherché à travers cette étude de cas est de valider la correspondance entre la représentation graphique des causalités et les indépendances. Le processus est explicité sur l'exemple du jardin de Holmes précédemment considéré.

5.3.1. Formalisation qualitative / quantitative

En premier lieu, nous procédons de la même manière que pour l'exemple précédent, et nous commençons par la construction intuitive de l'espace probabilisé à base de connaissances disponibles.

Avant de se faire, nous rappelons le contexte de l'exemple du jardin de Holmes :

« Ce matin là , alors que le temps est clair et sec, M Holmes sort de sa maison. Il s'aperçoit que la pelouse de son jardin est humide. (1) Il se demande alors s'il a plu pendant la nuit, ou s'il a simplement oublié de débrancher son arroseur automatique. Il jette alors un coup d'œil à la pelouse de son voisin, et s'aperçoit qu'elle est également humide.

Il en déduit alors (2) qu'il a probablement plu, et il décide de partir au travail sans vérifier sans arroseur automatique. »

D'après la description ci-dessus du problème, il est clair que l'ensemble des variables utiles regroupe quatre variables que nous pouvons noter A, P, J, V qui peuvent prendre chacune la valeur "vrai" ou "faux", où :

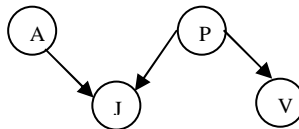
A : l'état de l'arroseur automatique

P : Il a plu

J : L'état de l'herbe du jardin de Holmes

V : L'état de l'herbe du jardin du voisin

Les relations causales entre ces variables est illustrée par le graphe suivant :



Formalisation probabiliste (quantitative)

Qualitativement ou de manière subjective, nous pouvons déduire l'ensemble des probabilités à priori ainsi que les probabilités de certains événements, soit marginales, soit conditionnellement à un autre événement. De même et d'après la description du problème, nous constatons que le fait qu'il ait plu cette nuit, et le fait que Holmes ait oublié de débrancher son arroseur automatique, sont deux événements complètement indépendants

▪ Probabilités à priori

| Evènement | Probabilité | |
|-----------|-------------|-------|
| A = V | 0.4 | Cas 1 |
| A = F | 0.6 | |
| P = V | 0.4 | Cas 2 |
| P = F | 0.6 | |

Les suppositions suivantes étaient à la base de l'attribution des probabilités résumées dans la table ci-dessus :

- ✓ Cas 1 : Holmes oublie assez souvent de débrancher son arroseur automatique.
- ✓ Cas 2 : La région est relativement pluvieuse.

▪ Probabilités conditionnelles

D'après le graphe causal modélisant l'ensemble des relations entre les quatre variables décrivant le contexte de l'exemple considéré, nous constatons que la valeur de J dépend des deux variables A et P. Cette dépendance est de type conditionnelle. Elle est illustrée à travers la table ci-dessous.

| | A = V | | A = F | |
|-------|-------|-------|-------|-------|
| | P = V | P = F | P = V | P = F |
| J = V | 1 | 1 | 1 | 0 |
| J = F | 0 | 0 | 0 | 1 |

De même, la valeur de la variable V dépend uniquement de la variable P . Donc, les probabilités correspondantes sont résumées sur la table suivante :

| | $P = V$ | $P = F$ |
|---------|---------|---------|
| $V = V$ | 1 | 0 |
| $V = F$ | 0 | 1 |

▪ Indépendances

Comme c'était mentionné juste en haut, les variables A et P sont indépendantes.

▪ Calculs et interprétations

A ce niveau nous cherchons à répondre aux deux questions posées dans l'exemple considéré. Nous détaillons, tout d'abord, le calcul, et son interprétation, effectué dans le but de répondre à la première question. Ensuite nous traitons la deuxième. Donc, nous distinguons deux traitements, chacun cherche la réponse à une question.

Traitement 1

La formalisation probabiliste de la question (1) posée dans l'exemple considéré est la suivante :

$$p(A = V / J = V)$$

et

$$p(P = V / J = V)$$

Calculons et comparons ces deux probabilités.

On a d'après la propriété d'inversion de Bayes (Bayes 1) :

$$p(A = V / J = V) = \frac{p(J = V / A = V) \cdot p(A = V)}{p(J = V)}$$

$$p(P = V / J = V) = \frac{p(J = V / P = V) \cdot p(P = V)}{p(J = V)}$$

et d'après le théorème de Bayes et d'indépendance de A et P , nous aurons :

$$\begin{aligned} p(J = V) &= p(J = V / A = V, P = V) \cdot p(A = V) \cdot p(P = V) + \\ &\quad p(J = V / A = V, P = F) \cdot p(A = V) \cdot p(P = F) + \\ &\quad p(J = V / A = F, P = V) \cdot p(A = F) \cdot p(P = V) + \\ &\quad p(J = V / A = F, P = F) \cdot p(A = F) \cdot p(P = F) \end{aligned}$$

D'où :

$$p(A = V / J = V) = 0.625$$

$$p(P = V / J = V) = 0.625$$

Les résultats obtenus de manière quantitative sont conformes aux connaissances intuitives induites de l'exemple dénonçant l'impossibilité de privilégier l'une des deux causes A et P suivant les seules informations disponibles.

Traitement 2 :

La formalisation probabiliste de la deuxième question, notée (2) dans la description de l'exemple est la suivante :

$$p(A = V/J = V, V = V)$$

et :

$$p(P = V/J = V, V = V)$$

Procédons maintenant au calcul et comparaison des deux probabilités ci-dessus.

Calculons tout d'abord :

$$p(P = V/J = V, V = V)$$

On a :

$$\begin{aligned} p(P = V / V = V) &= \frac{p(V = V / P = V) \cdot p(P = V)}{p(V = V)} \\ &= \frac{p(V = V / P = V) \cdot p(P = V)}{p(V = V / P = V) \cdot p(P = V) + p(V = V / P = F) \cdot p(P = F)} = 1 \end{aligned}$$

L'interprétation intuitive du résultat obtenu ($p(P = V / V = V) = 1$) est correcte, sachons que l'herbe du voisin est mouillée, donc il a certainement plu.

Et d'après la propriété probabiliste : si $p(A) = 1$, alors $p(A/B) = 1$, on aura :

$$p(P = V/J = V, V = V) = 1$$

Donc, il a certainement plu, est la conclusion finale atteinte.

De meme, en calculant $p(A = V/J = V, V = V)$ on aura :

$$p(A = V/J = V, V = V) = 0.4 = p(A = V)$$

Le fait que $p(A = V/J = V, V = V)$ n'est pas nulle est tout à fait logique : même si l'évènement "il a plu" est certain, ceci ne modifie pas la croyance *à priori* dans le fait que l'arroseur est resté branché.

5.4. Etude de cas N° 03

Cette dernière étude de cas a pour but de démontrer quantitativement la notion de circulation d'information, ou autrement dit, de prouver de manière probabiliste la circulation de l'information dans un graphe causal donné. Exactement, nous cherchons à valider quantitativement la correspondance entre l'indépendance et de circuits d'information. Pour se faire, nous considérons le même exemple de l'étude de cas N° 02 qui est celui du jardin de Holmes. Donc, la formalisation probabiliste du problème reste inchangée.

Nous savons que A et P sont intuitivement indépendants. Mais d'après le circuit $A \rightarrow J \leftarrow P$, et comme la relation est convergente en J et que J est connu, donc, l'information peut circuler de A à P. Donc, A et P ne sont pas indépendants conditionnellement à J. En ce qui suit, nous allons le démontrer quantitativement.

▪ Calcul

La formalisation probabiliste de la relation conditionnelle de A et P à J est la suivante :

$$p(A = F, P = F/J = V)$$

D'après la règle d'inversion de Bayes, on aura :

$$\begin{aligned} p(A = F, P = F / J = V) &= \frac{p(J = V / A = F, P = F) \cdot p(A = F, P = F)}{p(J = V)} \\ &= \frac{0 \times p(A = F, P = F)}{p(J = V)} \end{aligned}$$

$$= 0$$

Toutefois :

$$p(A = F/J = V) = 1 - p(A = V/J = V) = 0.375$$

et

$$p(P = F/J = V) = 1 - p(P = V/J = V) = 0.375$$

Donc :

$$p(A = F, P = F / J = V) \neq (p(A = F / J = V) \cdot p(P = F / J = V))$$

Cela veut dire que A et P ne sont pas indépendants conditionnellement à J. Ce qui confirme la conclusion atteinte qualitativement par application des règles de circulation de l'information sur un graphe causal.

▪ **Interprétation**

L'interprétation intuitive du résultat obtenu par calcul, est la suivante :

Si deux causes indépendantes déterminent un même effet connu, donc nécessairement l'une des deux lui a causé. Ce qui veut dire que les deux causes sont liées et elles ne sont plus indépendantes, *à posteriori*.

6. Références bibliographiques

Ben Mrad, A., Observations Probabilistes dans les Réseaux Bayésiens. Thèse de doctorat, Université de Sfax, 2015.

Feller, W., An Introduction to Probability Theory and Its Applications. Vol. 1, 3rd Edition, Wiley, 1968.

Leray, Ph., Réseaux bayésiens : introduction et apprentissage, modélisation et découverte de connaissances. 8èmes Journées Francophones Extraction et Gestion des Connaissances, Sophia Antipolis, 29 janvier 2008.

Naïm, P., Wuillemin, P. - H., Leray, P., Pourret, O., et Becker, A., Réseaux bayésiens. Groupe Eyrolles, 2^e édition, 2004.

Parent, O. et Eustache, J., Les Réseaux Bayésiens - A la recherche de la vérité. Cours Cognition et connaissance - Alain MILLE, Université Claude Bernard Lyon 1, 2006-2007.

Pearl, J., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (Morgan Kaufmann Series in Representation and Reasoning). Morgan Kaufmann, 1 edition, 1988.

Voir aussi :

Borzez, F., Mousseau, V., Pirlot, M., Siebert, X. and Viappiani, P. Bayesian Recommendations and Utility Elicitation with Intensity of Preference Statements, 2013.

Chateauneuf, A., Cohen, M. and Jaffray, J-Y. Decision under Uncertainty: The Classical Models. Decision-making Process: Concepts and Methods}, pp 385-400, ISTE/Wiley, 2009.

Fenton, N. and Neil, M. Making Decisions : Bayesian Nets and MCDA. Computer Science Department, London, 1999.

Jaffray, J-Y. Réseaux bayésiens. Concepts et méthodes pour l'aide à la décision à risque et incertain, pp 169-202, Hermès, Informatique et systèmes d'information, 2006.

Spanjaard, O. Contributions à la théorie de la décision algorithmique : optimisation multi-objectifs et optimisation dans l'incertain, Université Pierre et Marie Curie (UPMC), 2011.

Tonda. A., Lutton. E., Reuillon. R., Squillero. G. and Wullemmin. P-H. Bayesian network structure learning from limited datasets through graph evolution. 15th European Conference on Genetic Programming, pp 254-265, Springer-Verlag, 2012

Wiboonsak Watthayu and Yun Peng. A Bayesian network based framework for multi-criteria decision making. 17th International Conference on Multiple Criteria Decision Analysis, 2004.

Conclusion générale

Une entreprise exerce ses activités dans un environnement dynamique. Le milieu économique et social évolue, de nouvelles technologies émergent, le comportement des clients change, d'autres orientations stratégiques sont choisies par les actionnaires, des décisions internes sont prises à divers niveaux de l'entreprise ! Ainsi, nombreuses sont les origines de changements au sein d'une entreprise qui peuvent influencer son fonctionnement, ses flux, sa structure, sa position sur un marché, ou de façon générale la décision prise en son sein. Ceci représente la problématique clé pour l'informatique décisionnelle.

La matière présentée à travers le contenu de ce support de cours traite la modélisation en aide à la décision représentant la première étape de l'aide à la décision. Elle a pour but le fait d'évaluer les solutions pour aider à prendre une décision. Autrement dit, le modèle permet de connaître les conséquences des décisions avant de les appliquer. Le modèle est nécessaire à la prise de la décision, comme le plan est un préalable à la construction d'une maison. La modélisation permet alors de choisir la (ou les) meilleure(s) alternative(s) d'implantation, d'identifier les risques à gérer liés à ce changement, tout en bâtissant une vision commune du fonctionnement de l'entreprise.

Ce cours vise à développer les aptitudes des étudiants à élaborer et mettre en œuvre des modèles pertinents face à une situation de décision. A l'issue du cours, l'étudiant maîtrisera quelques méthodes de modélisation pour d'aide à la décision. Il saura les utiliser de façon opérationnelle dans le cadre de problèmes d'entreprises. Il aura aussi les éléments nécessaires pour prendre du recul et avoir un sens critique par rapport à ces méthodes, et ainsi en distinguer leurs performances et leurs limites d'application.

Annexe 1

Exercices

Exercice 1 :

Construisez l'arbre de décision correspondant à l'ensemble de règles suivantes :

1)

If MS > 5000 then Prêt = Yes

If MS ≤ 5000 and age ≤ 25 then Prêt = No

If MS ≤ 5000 and age > 25 and autres_comptes = Yes then Prêt = Yes

If MS ≤ 5000 and age > 25 and autres_comptes = No then Prêt = No

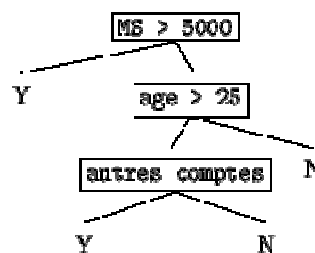
2)

If Enfant(E,P) and Masculin(P) then Père(P,E)

If Père(P,E1) and Père(P,E2) and Masculin(E1) then Frère(E1,E2)

Corrigé type :

1)



2) Impossible car il n'existe aucune liaison entre les deux règles.

Exercice 2 :

Considérons la tâche qui consiste à classer les conditions météorologiques en deux classes nommées P et N, selon les valeurs de quatre attributs qui décrivent respectivement la situation du ciel, la température, l'humidité et le vent.

14 exemples de classification sont donnés par le tableau suivant :

| NUM | CIEL | TEMP. | HUMI. | VENT | CLASSE |
|-----|------------|---------|---------|------|--------|
| 1 | ensoleillé | élevé | forte | non | N |
| 2 | ensoleillé | élevé | forte | oui | N |
| 3 | couvert | élevé | forte | non | P |
| 4 | pluvieux | moyenne | forte | non | P |
| 5 | pluvieux | basse | normale | non | P |
| 6 | pluvieux | basse | normale | oui | N |
| 7 | couvert | basse | normale | oui | P |
| 8 | ensoleillé | moyenne | forte | non | N |
| 9 | ensoleillé | basse | normale | non | P |
| 10 | pluvieux | moyenne | normale | non | P |
| 11 | ensoleillé | moyenne | normale | oui | P |
| 12 | couvert | moyenne | forte | oui | P |
| 13 | couvert | élevé | normale | non | P |
| 14 | pluvieux | moyenne | forte | oui | N |

- 1) Construisez des arbres de décision à partir de cette base d'exemple.
- 2) Quel est la différence entre les arbres construits ?
- 3) Dégager l'ensemble de règles de production exprimées par les arbres construits.

Corrigé type :

1) En cherchant à construire un arbre de décision qui classe correctement ces exemples, on remarque qu'un tel arbre n'est pas unique (voir figure 1 et figure 2).

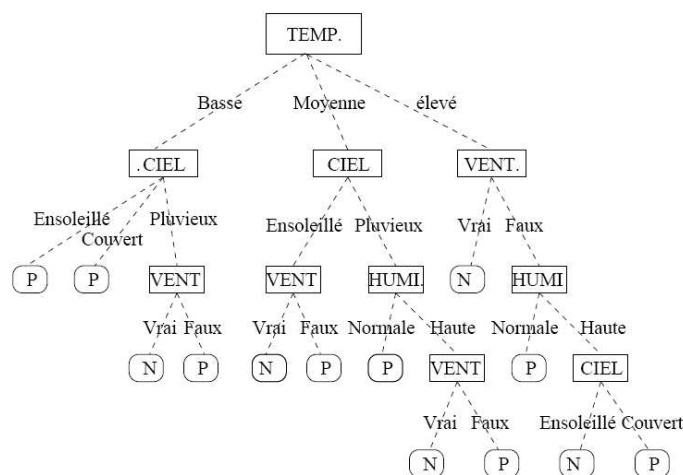


Figure 1. Arbre 1.

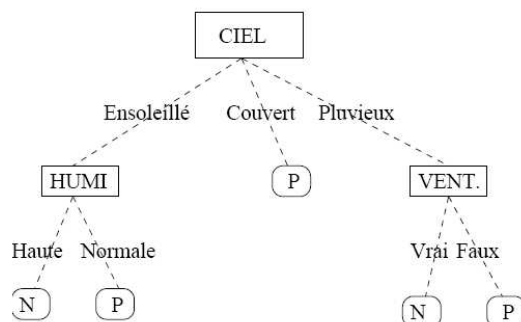


Figure 2. Arbre 2.

2) Chacun de ces arbres définit un algorithme avec des caractéristiques différentes comme par exemple :

- le nombre des nœuds,
- la longueur d'un chemin de la racine vers une feuille.

Puisque la différence de qualité est importante, il n'est pas question de se contenter d'un arbre quelconque construit au hasard.

3) Un arbre de décision exprime un ensemble de règles propositionnelles qui décrit les classes cibles et qui vérifie (classe correctement) tous les exemples inclus dans l'ensemble d'apprentissage.

L'arbre 2 exprime l'ensemble des règles suivantes :

- SI CIEL= Ensoleille & HUMI=Forte ALORS CLASSE=N
- SI CIEL= Ensoleille & HUMI=Normale ALORS CLASSE=P
- SI CIEL= Couvert ALORS CLASSE=P
- SI CIEL= Pluvieux & VENT=Oui ALORS CLASSE=N
- SI CIEL= Pluvieux & VENT=Non ALORS CLASSE=P

Autrement dit, le concept Classe=P est décrit par intention par l'hypothèse :

$(CIEL = Ensoleille \wedge HUMI = Normale) \vee (CIEL = Couvert) \vee (CIEL = Pluvieux \wedge VENT = Non)$

Exercice 3 :

Ahmed est propriétaire d'un bloc d'appartements au centre ville. Son seul objectif est de maximiser son profil. Chaque année, Ahmed détermine le loyer moyen qui chargera à ses locataires. C'est à partir de ce loyer moyen qu'il calcule le loyer spécifique de chaque appartement.

Ahmed fait en sorte que le nombre d'appartement occupées \times loyer moyen, soit une très bonne estimation du revenu total.

Q : Développer un DI qui calcule le profil en fonction du loyer moyen.

Hypothèses :

- Les dépenses fixes sont de 30000 DA par mois (assurance, entretien, taxes, etc).
- Les dépenses variables sont de 1500 DA par appartement occupé par mois.
- Il y a 50 appartements.
- La relation entre le loyer moyen et le taux d'occupation est donné à travers la table suivante :

| Loyer moyen (LM) | Taux d'occupation |
|-----------------------|-------------------|
| $4000 \leq LM < 4050$ | 95% |
| $4050 \leq LM < 4100$ | 93% |
| $4100 \leq LM < 4250$ | 90% |
| $4250 \leq LM < 4500$ | 85% |
| $4000 \leq LM$ | 75% |

Corrigé type :

Formules :

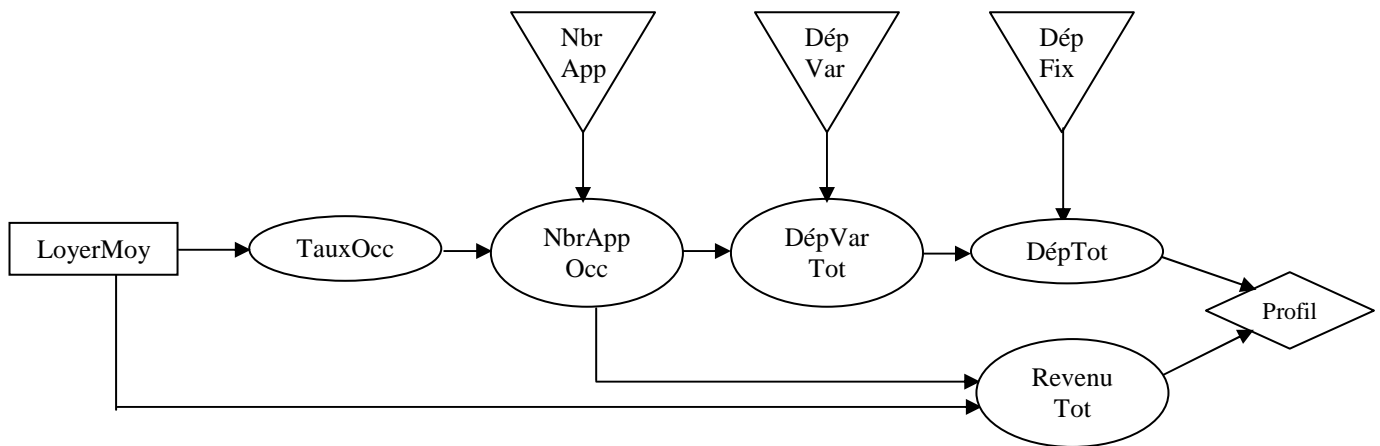
$$RevenuTot = LoyerMoy \times NbrAppOcc$$

$$DépTot = DépVarTot + DépFix$$

$$DépVarTot = DépVar \times NbrAppOcc$$

$$NbrAppOcc = \frac{TauxOcc \times NbrApp}{100}$$

$$Profil = RvenuTot - DépTot$$

**Exercice 4 :**

1) Représenter les règles logiques suivantes par un réseaux bayésien :

$$A \Rightarrow B$$

$$B \Leftrightarrow C$$

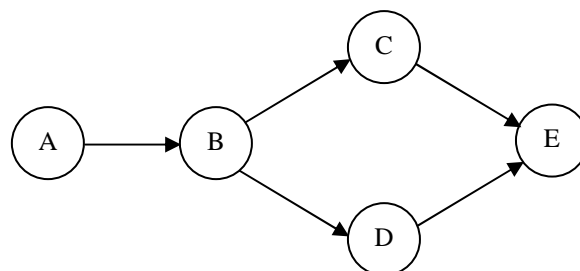
$$B \Leftrightarrow \overline{D}$$

$$E = C \text{ xor } D$$

2) Supposons que A soit faux, essayer de conclure logiquement sur E.

Corrigé type :

1)



2) Supposons que A soit faux, et essayons de conclure sur E.

Essayons d'abord le raisonnement logique.

- ✓ Comme A est faux, je ne peux pas utiliser la règle $A \Rightarrow B$ et donc je ne peux rien dire sur B. Toutefois, B est certainement soit vrai, soit faux.
- ✓ Supposons que B soit vrai. Dans ce cas, C est vrai, et D est faux, et E est donc vrai.
- ✓ Supposons que B soit faux. Dans ce cas, C est faux, et D est vrai, et E est donc vrai.

Donc : Si A est faux, E est vrai

Essayons maintenant la propagation locale des probabilités. Nous devons utiliser les tables de probabilités suivantes :

| | A = V | A = F |
|-------|-------|-------|
| B = V | 1 | 1/2 |
| B = F | 0 | 1/2 |

| | B = V | B = F |
|-------|-------|-------|
| C = V | 1 | 0 |
| C = F | 0 | 1 |

| | B = V | B = F |
|-------|-------|-------|
| D = V | 0 | 1 |
| D = F | 1 | 0 |

| | C = V | | C = F | |
|-------|-------|-------|-------|-------|
| | D = V | D = F | D = V | D = F |
| E = V | 0 | 1 | 1 | 0 |
| E = F | 1 | 0 | 0 | 1 |

Comme A est faux, la probabilité que B soit vrai (respectivement : faux) est $1/2$. Donc, la probabilité que C soit vrai est également $1/2$, et de même pour D.

Finalement, on conclut que la probabilité que E soit vrai est également de $1/2$.

Annexe 2

Examens 2009-2015

1. Examens de l'année 2009-2010

1.1. Examen ordinaire

Questions de cours : (6pts)

- 1) Dans le monde industriel, les principales étapes de la conception d'une mémoire d'entreprise sont calqués sur le schéma d'ingénierie d'un système à base de connaissances. Expliquer.
- 2) De quoi peut-on caractériser l'opération d'extraction des règles de production à partir d'un arbre de décision et l'opération inverse, c'est-à-dire l'opération de construction d'un arbre de décision à partir d'un ensemble de règles de production ?
- 3) Supposons $C1$ et $C2$ deux concepts d'une ontologie tel que : $C1$ est un sous concept de $C2$ ($C1 \subseteq C2$). Quelle est la relation qui peut exister entre l'extension de $C1$ et celle de $C2$, et entre l'intention de $C1$ et celle de $C2$?
- 4) Expliquer brièvement les différentes phases du processus de prise de décision. Quelle est la propriété de ce processus vis-à-vis du séquençement de ses phases ?
Est-ce que les clients peuvent avoir une influence sur cette propriété ? Justifier votre réponse.

Recommandation : Donner des réponses précises et concises.

Exercice 1 : (4pts)

On considère un espace de description comprenant les trois attributs *forme*, *taille* et *couleur* prenant respectivement les valeurs *rond* et *carré*, *petit* et *grand*, *bleu*, *blanc* et *rouge*. L'attribut cible est binaire de valeurs *oui* ou *non*.

Les données disponibles sont les suivantes (le " ? " correspond à une valeur manquante) :

| Forme | Taille | Couleur | Classe |
|-------|--------|---------|--------|
| rond | petit | bleu | oui |
| carré | grand | Rouge | non |
| rond | ? | blanc | oui |
| carré | petit | bleu | oui |
| rond | grand | bleu | oui |
| carré | grand | blanc | non |
| carré | ? | blanc | oui |
| carré | grand | bleu | non |
| carré | petit | rouge | oui |
| rond | grand | blanc | oui |

Ci-dessous, nous proposons deux méthodes pour déterminer les valeurs manquantes :

- **Méthode 1 (Valeur majoritaire de l'attribut) :** On remplace les valeurs manquantes par la valeur majoritaire prise par cet attribut sur l'échantillon complet.
- **Méthode 2 (Valeur majoritaire de l'attribut par classe) :** étant donné un exemple avec une valeur manquante, on remplace la valeur manquante par la valeur majoritaire prise par l'attribut correspondant pour les exemples de l'échantillon appartenant à la même classe.

- 1) D'après chacune des deux méthodes, quelle valeur associe-t-on sur notre échantillon ? Justifier votre réponse
- 2) Elaborer un arbre de décision modélisant cet échantillon.

Problème : (10pts)

Ahmed est le directeur d'un département informatique d'une moyenne entreprise et doit préparer un plan quinquennal d'acquisition de micro-ordinateur. La compagnie est dans une phase de croissance stable. Elle a présentement 90 employés et ce nombre augmente de 10% par année. En ce moment, 35% des employés sont équipés de micro-ordinateur, et la direction de la compagnie aimerait voir ce pourcentage augmenter à 80% d'ici cinq ans. Il prévoit aussi qu'à chaque année, il doit se débarrasser d'environ 15% des micro-ordinateurs que la compagnie possède à cause de désuétude ou de bris. De plus, aucun employé ne peut perdre son micro-ordinateur. Ahmed estime qu'il peut vendre tous ces vieux micro-ordinateurs pour environ 5000DA chacun et qu'un nouveau micro-ordinateur lui coûte 15000DA.

Ahmed croit que des pourcentages d'employés équipés de micro-ordinateurs de 45%, 52%, 62%, 70% et 80% à la fin de chacune des cinq prochaines années seraient raisonnables, mais il aimerait évaluer l'impact de ces valeurs sur son budget annuel d'acquisition pour les cinq prochaines années.

Questions :

- 1) Donner (Identifier) la liste des variables préliminaires et indiquer leur type.
- 2) Par un digramme d'influence, modéliser cette situation.
- 3) Donner la liste finale des variables et indiquer leur type.

Indication : Budget = Dépenses – Revenus

1.2. Examen de rattrapage

Questions de cours : (8pts)

Répondre par vrai ou faux en corrigeant les passages erronés :

- 1) Un SIAD est un système homme/machine.
- 2) Un décideur optimiste est un décideur neutre face au risque.
- 3) Les diagrammes d'influence montrent plus d'informations que les arbres de décision.
- 4) Il est très facile de convertir un arbre de décision en un DI.
- 5) Il est très facile de convertir un DI en un arbre de décision.
- 6) Les arbres de décision et les diagrammes d'influence sont des techniques concurrentes.
- 7) Les opérations de recherche et d'inférence de connaissances sont dépendantes du formalisme de représentation de connaissances.
- 8) Les buts d'une décision doivent être découverts avant le lancement du processus de la mise en œuvre de la décision.

Exercice 1 : (2pts)

1) Construisez l'arbre de décision correspondant à l'ensemble des deux règles suivantes :

If a and b then x
If c and d then x

2) Que constatez-vous ?

Problème : (10pts)

Un investisseur doit décider s'il achète un immeuble résidentiel, un édifice à bureaux ou un entrepôt. La rentabilité de son investissement sera déterminée par les futures conditions économiques (bonnes, moyennes ou mauvaises) qui représentent les états de la nature. Les profils correspondant à chaque décision et à chaque état de la nature (table des gains) sont évalués comme suit :

| Décision | Bonnes conditions économiques | Moyennes conditions économiques | Mauvaises conditions économiques |
|-------------|-------------------------------|---------------------------------|----------------------------------|
| Résidentiel | 80000 DA | 50000 DA | 20000 DA |
| Bureaux | 100000 DA | 60000 DA | -40000 DA |
| Entrepôt | 50000 DA | 30000 DA | 10000 DA |
| Aucun | 0 DA | 0 DA | 0 DA |

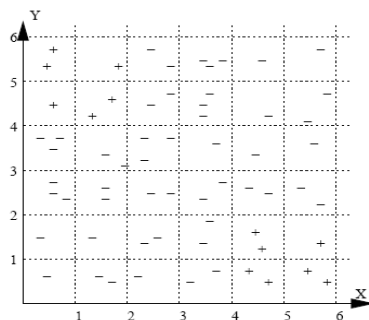
Les probabilités sont estimées pour chaque état du futur : l'économie sera bonne à 40%, moyenne à 20% et mauvaise à 40%.

1) Quel est l'investissement à faire selon chacun des critères suivants : Maximax, Maximin, Restriction de la perte (Seuil = 30000), Laplace ?

2) Modéliser le problème à l'aide d'un diagramme d'influence.

2. Examens de l'année 2010-2011**2.1. Examen ordinaire****Questions de cours : (9pts)**

- 1) De quoi peut-on caractériser la relation de spécialisation organisant les concepts d'une ontologie envers la symétrie, la réflexivité et la transitivité ?
- 2) Quel est l'objectif principal de l'opération de diffusion du processus d'ingénierie de connaissances ?
- 3) Les arbres de décision et les diagrammes d'influence sont des techniques complémentaires et non pas concurrentes. Expliquer.
- 4) Quelle technique utilise t-on pour évaluer un arbre de décision ? Comment l'applique t-on ?
- 5) Les buts d'une décision peuvent être découverts au moment même de l'exécution du processus de la mise en œuvre de la décision. Expliquer.
- 6) Quelles sont les opérations essentielles de manipulation de connaissances ? Est-ce qu'elles dépendent du formalisme de représentation de connaissances ?
- 7) Quelle est la différence entre un système interactif d'aide à la décision et un système décisionnel ?

Exercice 1 : (5pts)

1) Donner un arbre de décision qui classe correctement l'échantillon donné à travers la figure ci-dessus.

2) Etant donné les paramètres suivants :

a. de gauche à droite :

- Utilité_Occurrence1 (Résultat le plus répondu) = 35% ,
- Utilité_Occurrence2 (Résultat le plus répondu) = 20% ,
- Utilité_Occurrence3 (Résultat le plus répondu) = 10% ,
- Utilité_Occurrence4 (Résultat le plus répondu) = 41% ,
- Utilité_Occurrence5 (Résultat le plus répondu) = 27% .
- Utilité_Occurrence1 (Résultat le moins répondu) = 5% ,
- Utilité_Occurrence2 (Résultat le moins répondu) = 30% ,
- Utilité_Occurrence3 (Résultat le moins répondu) = 18% ,
- Utilité_Occurrence4 (Résultat le moins répondu) = 29% ,
- Utilité_Occurrence5 (Résultat le moins répondu) = 7% .

b. $P(\text{Résultat le plus répondu}) = 0.5$,
 $P(\text{Résultat le moins répondu}) = 0.5$.

Suivant les différents critères d'évaluation, pour quelle solution le décideur optera dans chacun des cas ?

Exercice 2 : (6pts)

Mohamed est le propriétaire d'un grand salon de coiffure pour homme au centre ville. Il a besoin d'un outil qui lui permettra de voir l'impact qu'aura sa décision quant au nombre d'employés à embaucher sur son profit hebdomadaire.

Le week-end représente la période de pointe de son salon. Durant cette période, il supervise ses employés. Durant le reste de la semaine, il travaille seul et il suffit à la tâche. Avec son expérience passée, il a pu évaluer la valeur des paramètres décrit à travers le tableau suivant :

| Paramètre | Description |
|-----------|--|
| NbrCltEmp | Nombre de clients servis par employé. Mohamed évalue que chaque employé peut servir 50 clients durant la période de pointe. Le revenu net par client, évalué à 150 DA. |
| RevNetClt | Cela correspond au prix du service moins le coût des marchandises. Naturellement, le salaire des employés n'est pas inclus dans le calcul de cette valeur. Le nombre de clients qui se présenteraient au salon de Mohamed durant la période de pointe. |
| NbrClt | Ce paramètre est évalué à 270 à ce temps-ci de l'année. Toutefois, il peut accroître comme il peut baisser, à d'autres moments de l'année. |
| Salaire | Le salaire, incluant toutes les charges sociales, est de 3000 DA par employé qui travaille durant la période de pointe. |

Développer un modèle qui permettra à Mohamed de calculer le profit hebdomadaire en fonction du nombre d'employés embauchés pour la période de pointe.

Remarque : Faites la différence entre la liste initiale de variables et la liste finale tout en indiquant le type de chaque variable.

2.2. Examen de rattrapage

Questions de cours : (bonne réponse +1, mauvaise réponse -0.5)

Répondre par vrai ou faux en corrigeant les passages erronés :

- 1) La plupart des formalismes de représentation des connaissances sont procédurales.
- 2) Le processus de prise de décision ne conduit pas linéairement de la sélection des buts à la conclusion.
- 3) Les étapes de la décision ont une signification temporelle.
- 4) Dans un DI, sauf les variables recevant plus qu'une flèche en entrée doivent être explicitées en formules mathématiques.
- 5) Les arbres de décision sont très précieux pour la phase de structuration et pour représenter de vastes problèmes, tandis que diagrammes d'influence fournissent plus de détails utiles pour passer à la phase d'analyse.
- 6) La restriction de la perte est le critère des décideurs pessimistes.

Exercice 1 : (4pts)

Soit l'algorithme suivant :

```

Algorithme EXO1
var
  X, Y, Z : entiers ;
DEBUT
  lire (X) ;
  lire (Y) ;
  lire (Z) ;
  Si (X<Y) and (Y<Z) alors
    écrire (X) ;
  Sinon
    Si (X<Y) alors
      Si (X<Z) alors
        écrire (X) ;
      Sinon
        écrire (Z) ;
      Fin Si
    Sinon
      Si (Y<Z) alors
        écrire (Y) ;
      Sinon
        écrire (Z) ;
      Fin si
    Fin Si
  Fin Si
FIN.

```

Question :

Construisez l'arbre de décision correspondant à l'algorithme EXO1.

Exercice 2 : (10pts)

Suivant l'état de certains paramètres (donnés ci-dessous), Mohammed pose la question suivante : ai-je les moyens d'avoir une voiture? Et quelle voiture ?!

- Caractéristiques physiques (modèle, couleur, confort, etc...).
- Prix
- Taux de financement
- Valeur de revente
- Frais d'utilisation (Assurances, entretien, consommation d'essence)

Plus précisément, il s'agit de décider vis-à-vis des points suivants :

- Quel modèle de voiture acheter. Le choix du modèle dictera le prix d'achat et le coût de la police d'assurance.
- s'il s'agit d'un engagement d'achat par faciliter, le modèle et par conséquent le prix d'achat, déterminera le taux d'intérêt qui déterminera à son tour le paiement annuel qui devra être effectué.
- la situation familiale au moment de l'engagement (célibataire ou marié) en plus du revenu mensuel de la personne impliquée, servent à fixer au préalable le taux d'intérêt.
- Les coûts d'assurance et le paiement annuel composent les coûts fixes annuels. Toutefois et de façon générale, d'autres coûts fixes qui ne dépendent pas du modèle de voiture choisi (par exemple, le coût d'installation des pneus d'hiver et d'été, et les lavages de voiture) peuvent être considérés.
- Les coûts fixes sont une des composantes du coût total. L'autre composante est celle correspondant au coût d'essence. Ce dernier dépend du prix de l'essence et du nombre de litres à consommer.
- Le nombre de litres dépend de la consommation du modèle de voiture choisi et du nombre de kilomètres que le propriétaire de véhicule prévoit faire au cours d'une année.
- Finalement, Mohamed aimerait aussi tenir compte d'autres facteurs qualitatifs dans sa prise de décision : est-ce qu'il va être plus populaire avec cette voiture? En d'autres termes, il faudra qu'il soit en mesure de donner une valeur *Image* pour chaque modèle de voiture.

Question :

Développez un modèle complet et détaillé aidant Mohamed pour choisir le modèle de voiture à acheter.

P.S : Le taux d'intérêt est une portion du prix total (pourcentage), que la personne cherchant à acheter une voiture par faciliter, peut payer chaque mois suivant sa situation familiale et son revenu mensuel.

3. Examens de l'année 2011-2012**3.1. Examen ordinaire****Questions de cours : (7pts = 1-1-1-1,5-1-1,5)**

- 1) Les stratégies progressives peuvent être exploitées pour prendre des décisions :
 1. Normalisées
 2. Non normalisées
 3. Structurées
 4. Non structurées
- 2) Les heuristiques ne peuvent pas être exploitées pour prendre des décisions :
 1. Normalisées
 2. Non normalisées
 3. Structurées

4. Non structurées

3) C'est quoi un modèle du décideur ?

4) Donner des exemples de tâches de pilotage de stratégie, de décision et des mesures.

5) Peut-on impliquer un système à base de connaissance dans un contexte décisionnel ? Expliquer.

6) Quelle est la différence entre un système à base de connaissance, un système à base de règles et un système expert ?

Exercice 1 : (6pts = 2,5-1,5-2)

Etant donné le cas d'un patient susceptible d'être atteint de la maladie m . À partir de cette seule connaissance, deux (02) stratégies sont à la disposition du médecin :

- **Stratégie $S1$** : consiste à ne rien donner au malade. Ne rien donner est dangereux si le patient est atteint de la maladie m (le coût du danger égale à 1000, soit utilité égale à -1000) et neutre s'il ne l'est pas ;
- **Stratégie $S2$** : consiste à donner le médicament M . Ce médicament n'est pas sans danger et peut entraîner des complications chez le patient. Suivant la situation, M sera plus moins bénéfique ou néfaste.

Le tableau suivant résume les utilités U_{ij} de ces deux stratégies pour les deux cas :

| | Maladie m | | Pas maladie m | |
|------|-------------------|-----------------|------------------|------------------|
| | $U_{1_1} = -1000$ | | $U_{1_2} = 0$ | |
| $S1$ | Pas complication | Complication | Pas complication | Complication |
| $S2$ | $U_{2_1} = 900$ | $U_{2_2} = 500$ | $U_{2_3} = -300$ | $U_{2_4} = -600$ |

1) Elaborer un arbre de décision modélisant la situation décrite ci-dessus.

2) Quelle est l'alternative recommandée pour des décideurs optimistes, pessimistes et ceux ayant un optimisme limité ?

3) Proposer des valeurs de probabilités pour rendre le problème un problème de prise de décision probabiliste, puis évaluer.

Exercice 2 : (7pts)

Mohamed est propriétaire d'un atelier de fabrication de pièces usinées. Les machines qu'il utilise présentement sont anciennes et il doit les remplacer, et il ne veut pas de machines différentes. Après une recherche auprès des différents fabricants de machines qu'il peut utiliser dans son atelier, Mohamed a résumé l'information qu'il a obtenue dans le tableau suivant :

| Type | Nom | Achat | Nombre d'employés | Coût par pièce | Capacité |
|------|----------------------|---------|-------------------|----------------|----------|
| 1 | Technotronics 2000 | 228 000 | 1 | 665 | 600 000 |
| 2 | Supertronics 2X | 279 000 | 2 | 627 | 800 000 |
| 3 | ACME | 175 000 | 1 | 775 | 450 000 |
| 4 | Master Technics 2001 | 272 000 | 2 | 643 | 700 000 |
| 5 | Super 10W30 | 280 000 | 2 | 641 | 750 000 |
| 6 | MultiTronic A1 | 304 000 | 3 | 587 | 800 000 |

Les caractéristiques de chaque machine sont :

| | |
|-------------------|---|
| Achat | Coût d'achat d'une machine. (On suppose qu'il n'y a pas de réduction à l'achat de plusieurs machines.) Ce coût représente la dépense annuelle nécessaire pour une machine et comprend, entre autre, les coûts d'emprunt, les assurances, etc. |
| Nombre d'employés | Le nombre d'employés nécessaires pour faire fonctionner une machine. En général, plus une machine est sophistiquée, plus elle coûte cher, plus elle a besoin d'employés et plus elle est performante. |
| Coût par pièce | Les coûts, autres que les matières premières, nécessaires pour produire une pièce. Ces coûts comprennent l'électricité, l'entretien, etc. |
| Capacité | Capacité de production annuelle d'une machine. |

Voici des informations supplémentaires :

- La capacité non-utilisée des machines n'a aucune valeur, mais les employés sont payés pour une année complète.
- La demande prévue pour l'an prochain est de 1400000 pièces et chaque pièce est vendue à 1200 DA.
- Le salaire annuel d'un opérateur de machine est 380000 DA.
- Le coût des matières premières pour fabriquer une pièce est 428 DA.

Question :

Développer un diagramme d'influence qui permettra à Mohamed de décider du type et du nombre de machines à acheter afin de maximiser son profit.

3.2. Examen de rattrapage

Questions de cours : (7pts = 2-3-2)

- 1) pourquoi les heuristiques ne peuvent pas être exploitées dans un contexte décisionnel normalisé ?
- 2) Donner des exemples de problèmes de structuration.
- 3) Répondre par vrai ou faux en corrigeant ce qu'il faut :
 - a. La largeur de l'arbre de décision détermine le nombre de séquences d'actions menant aux résultats.
 - b. La profondeur de l'arbre de décision détermine le nombre de séquences d'actions menant aux résultats.

Exercice 1 : (5pts = 1,5-2-1,5)

- 1) Construisez l'arbre de décision correspondant à l'ensemble de règles suivantes :

Règle 1: if Condition1 **and** Condition2 **then** Conclusion1

Règle 2: if Condition1 **and** Condition3 **and** Condition4 **then** Conclusion2

Règle 3: if Condition2 **and** Condition4 **then** Conclusion1

- 2) a. Donnez un nouvel ensemble de règles résultant de l'ensemble de départ en combinant les règles possibles.
- b. Construisez l'arbre de décision correspondant à ce nouvel ensemble.

Exercice 2 : (8pts)

Mohamed est propriétaire d'un salon de thé situé dans un centre commerciale. Ce salon de thé est ouvert pendant toute l'année. Il prévoit vendre 80000 pots de glace et, par expérience, il s'attend à une répartition des ventes selon les pourcentages suivants :

| Mois | Pourcentage |
|-----------|-------------|
| Janvier | 0.7% |
| Février | 1.1% |
| Mars | 1.8% |
| Avril | 3% |
| Mai | 10% |
| Juin | 20% |
| Juillet | 23% |
| Août | 26% |
| Septembre | 8% |
| Octobre | 3.2% |
| Novembre | 2.4% |
| Décembre | 0.8% |

Mohamed désire acheter une nouvelle machine à glace. Il a effectué des recherches auprès de quelques distributeurs locaux et ils lui ont proposé les machines suivantes :

| Type machine | Nombre de pots avant entretien | Coût d'entretien | Coût de fournitures par pot | Coût d'achat |
|--------------|--------------------------------|------------------|-----------------------------|--------------|
| 1 | 10 000 | 10000 DA | 7 DA | 200000 DA |
| 2 | 5 000 | 6000 DA | 12 DA | 180000 DA |
| 3 | 12 000 | 7500 DA | 5 DA | 150000 DA |

L'entretien doit être effectué à la fin du mois où le compteur dépasse le nombre de pots spécifié, et le compteur est alors remis à zéro.

En augmentant l'ensemble de contraintes de modélisation par :

- La fonction SOMME permettant de calculer implicitement la somme des valeurs d'une variable sur plusieurs itérations,
- Une structure de test (Si Condition alors Résultat 1 Sinon Résultat 2),

Développez un diagramme d'influence qui permettra de calculer les coûts totaux d'opération de la machine.

4. Examens de l'année 2012-2013

4.1. Examen ordinaire

Questions de cours : (8 pts)

- 1) Lesquels parmi les EIS, ESS, DSS et PSS ceux qui intègrent une phase d'analyse dans leur fonctionnement ? Expliquer.
- 2) Dans le contexte d'une entreprise, quels sont les éléments qui ont succité l'exploitation de l'informatique décisionnelle ?
- 3) Le temps et le changement ont été introduits dans la notion de processus de prise de décision en plus du choix. Expliquer.
- 4) pourquoi les heuristiques ne peuvent pas être exploitées dans un contexte décisionnel normalisé ?

Exercice 1 : (4 pts)

Des chercheurs psychologues ont obtenus les données suivantes concernant le fait qu'un étudiant haïsse un examen ou pas. Les attributs considérés sont : M, P et I.

| M | P | I | Hait examen ? |
|-----|-----|-----|---------------|
| oui | oui | oui | oui |
| oui | non | non | non |
| non | oui | oui | non |
| non | oui | non | oui |

1) Sur l'échantillon donné, construisez un arbre de décision parfait (optimal) pour prédire **Hait Examen ?**

2) Etant donné les paramètres suivants :

| | | | | | | | | |
|---------------------|-----|------|-----|-----|-----|-----|------|-----|
| Gains | 110 | 70 | 200 | 0 | 150 | 310 | 500 | 0 |
| Regrets | 80 | 110 | 50 | 200 | 65 | 10 | 4 | 310 |
| Probabilités | 0.2 | 0.25 | 0.3 | 0.6 | 0.4 | 0.9 | 0.16 | 0 |

Suivant les différents critères d'évaluation, pour quelle solution optera le décideur dans chacun des cas ?

Exercice 2 : (8 pts)

Mohamed est le responsable des activités scientifiques d'un laboratoire de recherche en Informatique. Durant la prochaine année, le laboratoire compte organiser quatre manifestations scientifiques auxquelles seront, bien sur, invités les membres du laboratoire ainsi que l'ensemble d'enseignants du département d'Informatique. Le tableau suivant présente la participation maximale prévue (nombre de participants) pour chacune des manifestations.

| Manifestation | Participation prévue |
|---------------|----------------------|
| CIIA | 170 |
| CVA | 350 |
| COSI | 280 |
| JCI | 410 |

Tableau 1

La participation est estimée pour un prix de 2000 DA. Des études de l'historique des manifestations scientifiques ont établi un facteur de variation de la participation en fonction des thèmes de recherche impliqués ainsi que des frais de participation selon le tableau suivant :

| Thèmes | Frais de participation | Facteur de variation |
|--|------------------------|----------------------|
| Security & Network Technologies | 1500 | 130 % |
| Information Technology | | |
| Computational Intelligence | | |
| Algorithmique discrète | 2500 | 85 % |
| Optimisation combinatoire | | |
| Représentation des connaissances et raisonnement | | |
| Systèmes d'information | 3000 | 50 % |
| Web services | | |
| Méthodes formelles, Sécurité des systèmes | | |
| Biométrie | 5000 | 30 % |
| Reconnaissance des formes | | |
| Cryptographie | | |

Tableau 2

Par exemple, si on demandait 1500 DA comme frais de participation avec comme thèmes : Security & Network Technologies, Information Technology, Computational Intelligence, alors la participation à COSI (manifestation 1) serait $280 \times 130 \% = 364$ participants.

Mohamed doit signer un contrat d'une année avec un propriétaire de salles de conférences. Il doit choisir dès maintenant trois salles dans lesquelles seront organisées toutes les manifestations. Les capacités et les prix des salles disponibles sont donnés dans le tableau suivant :

| Salle | Capacité | Loyer |
|-------|----------|--------|
| 1 | 1000 | 80000 |
| 2 | 500 | 40000 |
| 3 | 250 | 20000 |
| 4 | 1500 | 90000 |
| 5 | 2000 | 100000 |
| 6 | 380 | 35000 |

Tableau 3

Les revenus de ces manifestations consistent en les frais de participation. Les dépenses sont les frais de location des salles, les frais de préparation des proceeding (chaque participant doit disposer d'un proceeding dont le coût unitaire remonte à 1000 DA), les frais d'hébergement des participants (3500 DA pour chacun des participants), les frais de restauration (1000 DA pour chacun personne). De plus, 5 % des revenus seront récupérés par l'université.

Question :

Étant donné que les frais de participation aux différentes manifestations sont les mêmes, développer un diagramme d'influence qui permettra à Mohamed d'évaluer l'impact, sur le profit net total à partir du choix des salles et des frais de participation.

Remarque :

Tous les participants ont droit à l'hébergement qu'il soient résidents dans la même wilaya ou pas.

4.2. Examens de rattrapage

Questions de cours : (7pts)

1) Répondre par vrai ou faux en corrigeant ce qu'il faut :

Le nombre de règles dans l'ensemble minimal de règles de production dégagées à partir d'un arbre de décision égale au nombre de résultats.

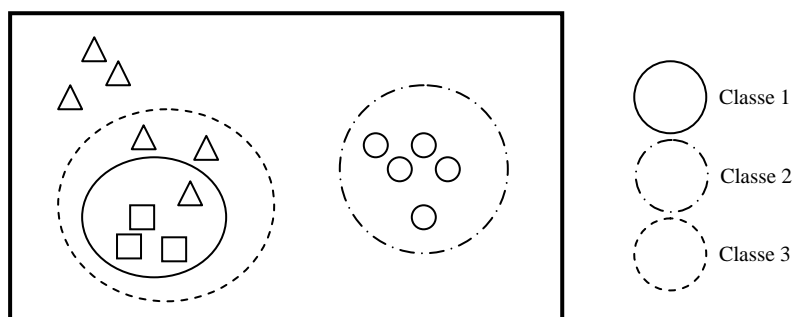
2) Dans quelles étapes du processus IDC intervient un EIS (respectivement, un ESS, un DSS et un PSS).

3) Quelle est la différence entre la modélisation dans le but d'aider à décider et celle appliquée dans un simple contexte informationnel (Système d'informations classique).

4) En distinguant les similarités et les différences entre les arbres de décision et les diagrammes d'influence, faites une comparaison entre eux qui s'articule autour des points critiques des deux techniques.

Exercice 1 : (6pts)

Soit la figure ci-dessous :



Questions :

- 1) Quelle est la nature (le type) du problème traité à travers cette figure ?
- 2) Est-il possible de modéliser cette situation par un arbre de décision ? Si oui, élaborer l'arbre correspondant, sinon, préciser les choses empêchant la modélisation puis proposer une solution et modéliser cette nouvelle situation par un arbre de décision.

Exercice 2 : (7pts)

Mohamed compte d'ouvrir un atelier de couture. Il doit décider du nombre d'employés à embaucher en plus du nombre et du type de machines à acheter. Après une recherche auprès des différents fabricants de machines qu'il peut utiliser dans son atelier, Mohamed a résumé l'information qu'il a obtenue dans le tableau suivant :

| Type | Nom | Achat (DA) | Coût par Pièce (DA) | Capacité |
|------|------------|------------|---------------------|----------|
| 1 | Sinjer | 124000 | 3400 | 600 000 |
| 2 | Toyota | 178000 | 2800 | 700 000 |
| 3 | Lock | 87000 | 1000 | 450 000 |
| 4 | Elna | 90600 | 1200 | 500 000 |
| 5 | Janome | 95000 | 1400 | 550 000 |
| 6 | Lewenstein | 104000 | 2100 | 650 000 |

Où :

- Achat : représente le coût d'achat d'une machine. Ce coût représente la dépense annuelle nécessaire pour une machine et comprend, entre autre, les coûts d'emprunt, les assurances, etc.
- Coût par pièce : représente les coûts, autres que les matières premières, nécessaires pour produire une pièce. Ces coûts comprennent l'électricité, l'entretien, etc.
- Capacité : représente la capacité de production annuelle d'une machine.

Voici des informations supplémentaires :

- La demande prévue pour la première année après l'ouverture de l'atelier est de 2000000 pièces (tabliers), chacune sera vendue à 1200 DA.
- Le coût des matières premières pour fabriquer une pièce (un tablier) est 500 DA.

Mohamed prévoit d'embaucher deux (02) catégories d'employés, ceux qui ont une bonne expérience dans le domaine (couturiers qualifiés) et d'autres pour accomplir de simples tâches (placement de boutons, ...). Le salaire annuel d'un couturier qualifié est de 300000 DA et celui d'un simple employé est de 204000 DA.

Question :

Développer un diagramme d'influence qui permettra à Mohamed de décider du type et du nombre de machines à acheter ainsi que du nombre d'employés à embaucher afin de maximiser son profit.

5. Examens de l'année 2013-2014**5.1. Examen ordinaire****Questions de cours : (7 pts)**

- 1) Un modèle de décideur devrait contenir également un modèle de l'environnement, même si ce dernier n'est pas sophistiqué. Expliquer.
- 2) Classer les exemples de décisions suivant les différents types de décisions :
 - a- Gestion d'un département,
 - b- le job à entreprendre par un nouveau bachelier une fois qu'il terminera ses études,
 - c- Fixer les prix des produits d'une entreprise suivant les changements continuels de l'état du marché,

d- La mise en place d'un système d'évaluation de la qualité de produits,

e- La détermination de l'emplacement d'une usine.

3) Y'a-t-il une relation entre les décisions structurées et celles normalisées ? Expliquer.

4) Quels sont les points qui déterminent qu'un diagramme d'influence soit correcte ou faux sans l'analyser de manière détaillée ?

Exercice 1 : (6 pts)

Etant donnée la base d'exemples suivantes :

| Nº Exemple | Maladie | Durée Maladie (mois) | Dettes (DA) | Profession | Etat |
|------------|---------|----------------------|-------------|------------|--------------|
| 1 | Aucune | 0 | 0 | Pilote | Très bon |
| 2 | Grippe | 2 | 50000 | Plombier | Bon |
| 3 | Concert | 18 | 2000000 | Aucune | Très mauvais |
| 4 | Grippe | 1 | 6000 | Aucune | Mauvais |
| 5 | SIDA | 7 | 0 | Plombier | Très mauvais |
| 6 | Concert | 6 | 0 | Médecin | Mauvais |
| 7 | Aucune | 0 | 0 | Pilote | Très Bon |
| 8 | Concert | 22 | 0 | Pilote | Très mauvais |
| 9 | Aucune | 0 | 0 | Plombier | Très bon |
| 10 | Aucune | 0 | 2500000 | Médecin | Mauvais |

1) Modéliser cette base par un arbre de décision.

2) Dégager l'ensemble des règles de production.

3) Optimiser l'ensemble des règles.

Exercice 2 : (7 pts)

Mohamed est le directeur d'un théâtre. Durant le prochain semestre, douze pièces théâtrales auront lieu dans ce théâtre. Deux pièces théâtrales seront organisées en un même mois. Les deux premières semaines d'un mois seront consacrées à la préparation d'une même pièce pour enfin la présenter durant le premier jour du week-end de la deuxième semaine. Le deuxième jour du week-end de cette dernière semaine sera consacré à la préparation du théâtre pour la prochaine exposition (entretien, décoration, ...). Des amis de(s) artiste(s) présentant une certaine pièce théâtrale seront invités le jour de sa présentation à condition que ce nombre ne dépassera pas 5% de la capacité de la salle d'exposition. Le tableau suivant présente le nombre d'entrées prévues pour chacune des pièces ainsi que le cachet qui sera versé aux artistes à la fin d'une exposition.

| Mois | Pièce | Nbr d'entrées | Nbr invités | Cachet (DA) |
|------|-------|---------------|-------------|-------------|
| 1 | 1 | 1700 | 60 | 50000 |
| | 2 | 1620 | 45 | 47000 |
| 2 | 1 | 1500 | 50 | 45000 |
| | 2 | 2000 | 70 | 52000 |
| 3 | 1 | 2500 | 165 | 70000 |
| | 2 | 1000 | 30 | 30000 |
| 4 | 1 | 1200 | 45 | 36000 |
| | 2 | 2200 | 90 | 55000 |
| 5 | 1 | 1000 | 60 | 28000 |
| | 2 | 2500 | 120 | 60000 |
| 6 | 1 | 1800 | 90 | 52000 |
| | 2 | 1500 | 10 | 46000 |

- Table 1 -

Le prix d'entrée est estimé à 500 DA. Le tableau suivant présente un facteur de variation du nombre d'entrées en fonction du prix d'entrée.

| Prix d'entrée (DA) | Facteur de variation | |
|-----------------------|-------------------------|-------------|
| 200 | 300 % | - Table 2 - |
| 350 | 150 % | |
| 600 | 75 % | |
| 800 | 30 % | |

Par exemple, si on demandait 200 DA alors le nombre d'entrées lors de l'exposition de la première pièce théâtrale du premier mois serait $1700 \times 300 \% = 510000$ spectateurs. Les revenus seront versés à la caisse du théâtre.

Le tableau suivant résume d'autres paramètres que Mohamed utilise :

| Paramètre | Description | |
|----------------|---|-------------|
| FraisPayement | <i>Les frais de payement du personnel</i> , incluant toutes les charges sociales. Il est évalué à 150000 DA par mois. | - Table 3 - |
| FraisEntretien | Des entretiens de la salle d'exposition auront lieu à la fin du mois. <i>Les frais d'entretien</i> remontent à 30000 DA. | |
| CapacitéSalle | La salle d'exposition du théâtre peut accueillir jusqu'à 2500 spectateurs. | |

Question :

Étant donné que les frais d'entrées sont les mêmes pour toutes les expositions, développer un diagramme d'influence qui permettra à Mohamed d'évaluer l'impact, sur le profit net total à partir des frais d'entrées.

5.2. Examens de rattrapage

Questions de cours : (6pts)

1) Répondre par vrai ou faux en corrigeant ce qu'il faut : (*Bonne réponse +1, mauvaise réponse - 0.5*)

- a- La notion de décideur isolé est indispensable.
- b- les étapes de la décision ont une signification logique et temporelle.
- c- Toute décision normalisée est structurée.
- d- Le processus de prise de décision ne conduit pas linéairement de la sélection des buts à la conclusion.

2) Expliquer la dernière phase du modèle IDC.

Exercice 1 : (4pts)

Soit l'algorithme suivant :

```

    Algorithme Petit ;
var
    X, Y, Z : entiers ;
DEBUT
    lire (X) ;
    lire (Y) ;
    lire (Z) ;
    Si (X<Y) and (Y<Z) alors
        écrire (X) ;
    Sinon
        Si (X<Y) alors
            Si (X<Z) alors
                écrire (X) ;
            Sinon
                écrire (Z) ;
        Fin Si
    Sinon
        Si (Y<Z) alors
            écrire (Y) ;

```

Sinon
 écrire (Z) ;
Fin si
Fin Si
Fin Si

FIN.

Question :

Construisez l'arbre de décision correspondant à l'algorithme *Petit*.

Exercice 2 : (10pts)

Suivant l'état de certains paramètres (donnés ci-dessous), Mohamed pose la question suivante: ai-je les moyens d'avoir une voiture? Et quelle voiture ?!

- Caractéristiques physiques (modèle, couleur, confort, etc...).
- Prix
- Taux de financement
- Valeur de revente
- Frais d'utilisation (Assurances, entretien, consommation d'essence)

Plus précisément, il s'agit de décider vis-à-vis des points suivants :

- Quel modèle de voiture acheter. Le choix du modèle dictera le prix d'achat et le coût de la police d'assurance.
- s'il s'agit d'un engagement d'achat par faciliter, le modèle et par conséquent le prix d'achat, déterminera le taux d'intérêt qui déterminera à son tour le paiement annuel qui devra être effectué.
- la situation familiale au moment de l'engagement (célibataire ou marié) en plus du revenu mensuel de la personne impliquée, servent à fixer au préalable le taux d'intérêt.
- Les coûts d'assurance et le paiement annuel composent les coûts fixes annuels. Toutefois et de façon générale, d'autres coûts fixes qui ne dépendent pas du modèle de voiture choisi (par exemple, le coût d'installation des pneus d'hiver et d'été, et les lavages de voiture) peuvent être considérés.
- Les coûts fixes sont une des composantes du coût total. L'autre composante est celle correspondant au coût d'essence. Ce dernier dépend du prix de l'essence et du nombre de litres à consommer.
- Le nombre de litres dépend de la consommation du modèle de voiture choisi et du nombre de kilomètres que le propriétaire de véhicule prévoit faire au cours d'une année.
- Finalement, Mohamed aimerait aussi tenir compte d'autres facteurs qualitatifs dans sa prise de décision : est-ce qu'il va être plus populaire avec cette voiture? En d'autres termes, il faudra qu'il soit en mesure de donner une valeur *Image* pour chaque modèle de voiture.

Question :

Développez un modèle complet et détaillé aidant Mohamed pour choisir le modèle de voiture à acheter.

P.S : Le taux d'intérêt est une portion du prix total (pourcentage), que la personne cherchant à acheter une voiture par faciliter, peut payer chaque mois suivant sa situation familiale et son revenu mensuel.

6. Examens de l'année 2014-2015

6.1. Examen ordinaire

Questions de cours : (2+1+2 = 5pts)

- 1) De nombreux problèmes dans les organisations peuvent s'analyser en terme d'allocation de ressources où un décideur doit allouer des ressources peu abondantes pour des activités variées afin d'optimiser un objectif mesurable. Expliquer.
- 2) Positionner les outils d'aide à la décision classés suivant le niveau de décision sur le modèle IDC.
- 3) Répondre par vrai ou faux tout en expliquant votre réponse :
 - a- Tout comme les réseaux bayésiens, les arbres de décision offrent une aide qualitative et quantitative.
 - b- Tout comme les réseaux bayésiens, les diagrammes d'influence offrent une aide qualitative et quantitative.

Exercice 1 : (8 pts)

1) En donnant son cours pour les étudiants de la 2ème année médecine, Mohamed qui est un spécialiste ORL et maître assistant à la faculté de médecine d'Alger, Mohamed signale les notes ci-dessous pour diagnostiquer les quatre maladies suivantes:

- 1- Migraine : maux (douleurs) de tête,
- 2- Rhume : maux de gorge accompagnés de fièvre / toux accompagnée de fièvre avec absence de douleurs,
- 3- Maux de gorge : maux de gorge avec absence de fièvre,
- 4- Refroidissement : toux avec absence de fièvre ou de douleurs,
- 5- Aucune des quatre maladies : absence de douleurs, de toux ou de fièvre.

Q1) Suivant le type du problème traité, faites extraire une base d'exemples non triviale.

Q2) Modéliser la base d'exemple obtenue par un arbre de décision.

2) Etant donné les paramètres suivants :

| | | | | | | | | | | |
|----------------------------------|-----|-----|-----|-----|----|----|---|----|-----|----|
| Gain_{médic} Max | 200 | 110 | 250 | 150 | 80 | 70 | 0 | 50 | 140 | 75 |
| Gain_{médic} réel | 110 | 100 | 200 | 120 | 65 | 90 | 0 | 45 | 85 | 0 |

Q) Suivant les différents critères d'évaluation, pour quelle solution optera le décideur dans chacun des cas ?

P.S : Gain_{médic} est le gain espéré en prenant un médicament en étant malade d'une certaine maladie.

Exercice 2 : (7 pts)

Pour un devoir à la maison donné par l'enseignant de la matière "sciences naturelles", Meriem, élève en 2ème année moyenne, a mené une petite recherche sur les symptômes de certaines maladies.

En cherchant sur Internet, en lisant des revues médicales et en assistant à des programmes télévisés parlant du sujet, Meriem a pu noter les points suivants :

- Une tuberculose peut être captée en voyageant en certaines régions asiatiques.
- Une exposition aux produits toxiques ou le tabagisme (être fumeur) peuvent causer un cancer de poumons qui peut entraîner une dyspnée de son tour.
- La grande partie des fumeurs sont des hommes plutôt que de femmes, et adolescents ou adultes.
- Le tabagisme peut entraîner une bronchite qui peut se développer en dyspnée.
- la radiologie confirme le fait d'être atteint d'une tuberculose ou d'un cancer de poumons.

- 1) Modéliser la situation ci-dessus par un réseau bayésien.
- 2) Formuler en termes de probabilités :
 - a- la probabilité qu'une vieille femme ait un cancer malin.
 - b- la probabilité qu'un jeune adolescent souffrant d'une forte dyspnée ait un cancer malin.
- 3) Etant donnée une personne fumeuse et une probabilité $P(X,Y)$ qui suit le tableau suivant :

| $P(X,Y)$ | $X_1 = \text{masculin}$ | $X_2 = \text{féminin}$ |
|---------------------------|-------------------------|------------------------|
| $Y_1 = \text{enfant}$ | 0.0015 | 0.0003 |
| $Y_2 = \text{adolescent}$ | 0.14 | 0.071 |
| $Y_3 = \text{adulte}$ | 0.7 | 0.26 |
| $Y_4 = \text{vieux}$ | 0.1 | 0.04 |

Q) Calculer la probabilité que la personne soit un homme (mâle).

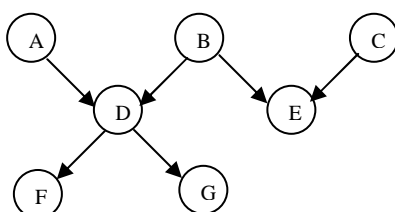
6.2. Examen de rattrapage

Questions de cours : (7 pts)

- 1) Répondre par vrai ou faux : (*Bonne réponse +0.5, mauvaise réponse - 0.25*)
 - a- Les stratégies progressives peuvent être exploitées pour prendre des décisions normalisées.
 - b- Les stratégies progressives peuvent être exploitées pour prendre des décisions structurées.
 - c- Les stratégies progressives peuvent être exploitées pour prendre des décisions non normalisées.
 - d- Les stratégies progressives peuvent être exploitées pour prendre des décisions non structurées.
 - e- Les heuristiques ne peuvent pas être exploitées pour prendre des décisions normalisées.
 - f- Les heuristiques peuvent être exploitées pour prendre des décisions structurées.
 - g- Les heuristiques peuvent être exploitées pour prendre des décisions non normalisées.
 - h- Les heuristiques ne peuvent pas être exploitées pour prendre des décisions non structurées.
 - i- Le nombre de règles dans l'ensemble minimal de règles de production dégagées à partir d'un arbre de décision égale au nombre de résultats.
- 2) Lesquels parmi les EIS, ESS, DSS et PSS ceux qui intègrent une phase d'analyse dans leur fonctionnement ? Expliquer.
- 3) La modélisation qui désire savoir comment un décideur prend une décision doit s'intéresser à la façon dont le décideur modélise le monde. Expliquer puis donner un exemple reflétant le passage souligné.
- 4) Quel est le type de situation (décision) modélisée par un arbre de décision trop profond et trop large ? Expliquer.

Exercice 1 : (4 pts)

Etant donné le graphe causal ci-dessous, quels sont les nœuds indépendants l'un de l'autre ?



Exercice 2 : (8 pts)

Mohamed est propriétaire du centre commerciale de Bab Ezzouar situé à Alger. Le centre est ouvert durant toute l'année où la répartition du nombre de visiteurs varie suivant les saisons résumées dans le tableau 1.

| Saisons | Taux Visiteur Saison | |
|-----------|----------------------|------|
| | Jour | Soir |
| Hivers | 8% | 2% |
| Printemps | 17% | 8% |
| Eté | 20% | 30% |
| Automne | 9% | 6% |

- Tableau 1 -

Le mois de Ramadhan représente une période spéciale de l'année, et vue que le centre contient beaucoup d'équipements de loisir et d'espaces familiaux, le nombre de visiteurs augmente en fonction de la saison suivant les détails illustrées à travers le tableau 2.

| Saison | Augmentation | |
|-----------|--------------|------|
| | jour | soir |
| Hivers | 15% | 3% |
| Printemps | 20% | 22% |
| Eté | 16% | 50% |
| Automne | 17% | 8% |

- Tableau 2 -

Les frais de parking représentent une portion des revenus du centre. Ils sont estimés à 200 DA par place (véhicule) durant toute l'année, sauf pour le mois de Ramadhan où ils sont réduits à 150 DA par place le jour et à 100 DA le soir.

Le tableau suivant (tableau 3) résume d'autres paramètres que Mohamed utilise :

| Paramètre | Description |
|-------------|---|
| NbrVisitTot | Nombre total de visiteur durant toute l'année, estimé à 3000000 |
| NbrVéhicule | 40% des visiteurs sont véhiculés |

- Tableau 3 -

Question :

En augmentant l'ensemble de contraintes de modélisation par :

- Une structure de test (Si Condition alors Résultat 1 Sinon Résultat 2),

développer un diagramme d'influence modélisant le contexte présenté ci-dessus tout en permettant à Mohamed d'estimer les revenus du centre par saison en lui explicitant les paramètres gestionnaires du jour de ceux du soir.