

Chapitre 7

Mise en œuvre des réseaux bayésiens

Nous abordons maintenant la mise en œuvre des réseaux bayésiens dans des applications pratiques. Dans ce chapitre, nous présentons essentiellement des aspects méthodologiques, en essayant de répondre aux trois questions suivantes : pourquoi, où (dans quelles applications) et comment utiliser des réseaux bayésiens ?

Les chapitres suivants seront consacrés, d'une part à une revue générale d'applications dans le monde, et d'autre part à quatre études de cas détaillées.

7.1 Pourquoi utiliser des réseaux bayésiens ?

Selon le type d'application, l'utilisation pratique d'un réseau bayésien peut être envisagée au même titre que celle d'autres modèles : réseau de neurones, système expert, arbre de décision, modèle d'analyse de données (régression linéaire), arbre de défaillances, modèle logique. Naturellement, le choix de la méthode fait intervenir différents critères, comme la facilité, le coût et le délai de mise en œuvre d'une solution. En dehors de toute considération théorique, les aspects suivants des réseaux bayésiens les rendent,

dans de nombreux cas, préférables à d'autres modèles :

- ① **Acquisition des connaissances.** La possibilité de rassembler et de fusionner des connaissances de diverses natures dans un même modèle : retour d'expérience (données historiques ou empiriques), expertise (exprimée sous forme de règles logiques, d'équations, de statistiques ou de probabilités subjectives), observations. Dans le monde industriel, par exemple, chacune de ces sources d'information, quoique présente, est souvent insuffisante individuellement pour fournir une représentation précise et réaliste du système analysé.
- ② **Représentation des connaissances.** La représentation graphique d'un réseau bayésien est explicite, intuitive et compréhensible par un non-spécialiste, ce qui facilite à la fois la validation du modèle, ses évolutions éventuelles et surtout son utilisation. Typiquement, un décideur est beaucoup plus enclin à s'appuyer sur un modèle dont il comprend le fonctionnement qu'à faire confiance à une boîte noire.
- ③ **Utilisation des connaissances.** Un réseau bayésien est polyvalent : on peut se servir du même modèle pour évaluer, prévoir, diagnostiquer, ou optimiser des décisions, ce qui contribue à rentabiliser l'effort de construction du réseau bayésien.
- ④ **Qualité de l'offre en matière de logiciels.** Il existe aujourd'hui de nombreux logiciels pour saisir et traiter des réseaux bayésiens. Ces outils présentent des fonctionnalités plus ou moins évoluées : apprentissage des probabilités, apprentissage de la structure du réseau bayésien, possibilité d'intégrer des variables continues, des variables d'utilité et de décision, etc.

Nous allons à présent étudier plus en détail ces différents aspects de l'utilisation de réseaux bayésiens.

7.1.1 Acquisition des connaissances

► Un recueil d'expertise facilité

Comme nous l'avons vu dans le chapitre 1 page 3, la représentation des connaissances utilisées dans les réseaux bayésiens est la plus intuitive possible : elle consiste simplement à relier des causes et des effets par des flèches. Pratiquement toute représentation graphique d'un domaine de connaissances peut être présentée sous cette forme.

De nombreuses expériences montrent qu'il est souvent plus facile pour un expert de formaliser ses connaissances sous forme de graphe causal que sous forme de système à base de règles, en particulier parce que la formulation de règles sous la forme SI... ALORS est très contraignante, et peut être facilement mise en défaut.

Certains auteurs considèrent qu'il existe une différence de nature entre les deux processus d'acquisition de connaissances. Lorsqu'on essaie de mettre au point un système expert, par exemple pour une application de diagnostic, l'expert doit décrire le processus de raisonnement qui le conduit de ses observations à une conclusion. En revanche, un modèle fondé sur un graphe causal décrit la perception de l'expert du fonctionnement du système. Effectuer un diagnostic n'est alors qu'une résultante de cette modélisation.

► Un ensemble complet de méthodes d'apprentissage

Comme nous l'avons abordé dans la première partie, et détaillé dans la partie théorique, les algorithmes actuels permettent d'envisager l'apprentissage de façon très complète :

- En l'absence totale de connaissances, on peut rechercher à la fois la structure du réseau la plus adaptée, c'est-à-dire les relations de dépendance et d'indépendance entre les différentes variables, et les paramètres, ou probabilités, c'est-à-dire la quantification de ces relations.
- Si l'on dispose de connaissances *a priori* sur la structure des causalités, et d'une base d'exemples représentative, la détermination des matrices de probabilités conditionnelles, qui sont les paramètres du réseau, peut être effectuée par simple calcul de fréquences, par détermination du *maximum de vraisemblance*, ou par des méthodes bayésiennes.

Ces méthodes peuvent être étendues dans le cadre de bases de données incomplètes. Dans l'optique de rechercher un compromis entre apprentissage et généralisation, il est également possible d'effectuer des apprentissages en contraignant la structure du réseau.

► Un apprentissage incrémental

Le principe général de l'apprentissage dans les réseaux bayésiens est décrit par la formule générale :

$$A_{\text{Posteriori}} \propto \text{Vraisemblance} \times A_{\text{Priori}}$$

Cette formule, que nous avons établie dans la partie théorique, conditionne la modification de la connaissance contenue dans le réseau par l'acquisition de nouveaux exemples. Elle s'interprète en disant que la connaissance contenue *a priori*, ou à un instant quelconque, dans le réseau, est transformée *a posteriori* en fonction de la vraisemblance de l'observation

des exemples étudiés selon la connaissance initiale. Autrement dit, plus les exemples observés s'écartent de la connaissance contenue dans le réseau, plus il faut modifier celle-ci.

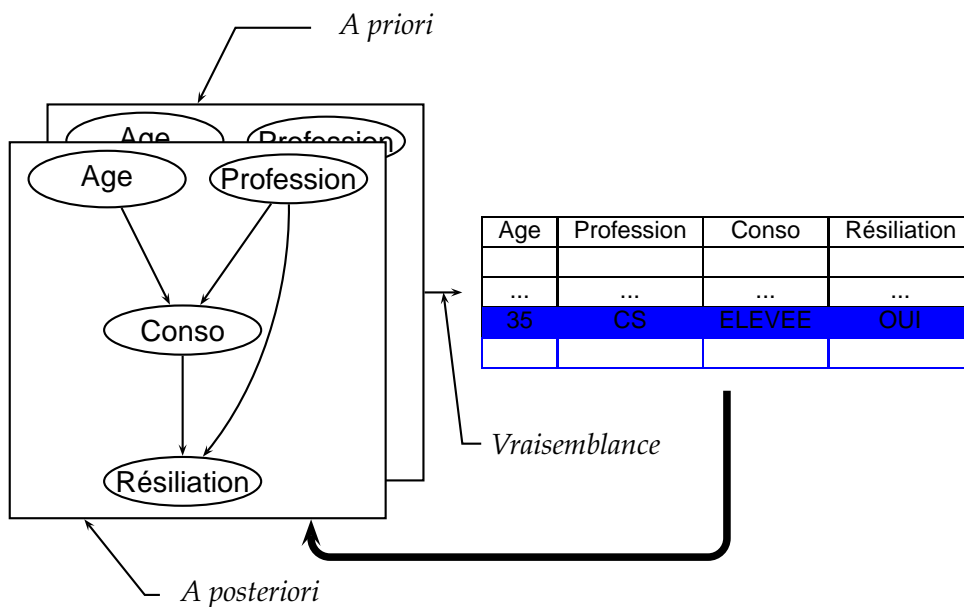


FIG. 7.1 Un exemple d'apprentissage incrémental (data mining)

Théoriquement, cette formule, qui n'est autre que la formule de Bayes appliquée à la connaissance, est valable aussi bien pour l'apprentissage de paramètres que pour l'apprentissage de structure. Aucune des techniques concurrentes, ni les réseaux neuronaux, ni les arbres de décision, ne permet de prendre en compte ce problème de la mise à jour des modèles de connaissance de façon aussi naturelle, même si aujourd'hui sa mise en œuvre dans les réseaux bayésiens n'est possible techniquement que dans certains cas particuliers.

Nous pensons que la capacité d'apprentissage incrémental est essentielle, car elle autorise l'évolution des modèles. Toute démarche de modélisation qui ne concerne pas les sciences de la nature doit intégrer les évolutions de l'environnement modélisé, et donc faire dépendre le modèle du temps. L'apprentissage incrémental est une réponse possible à ce problème.

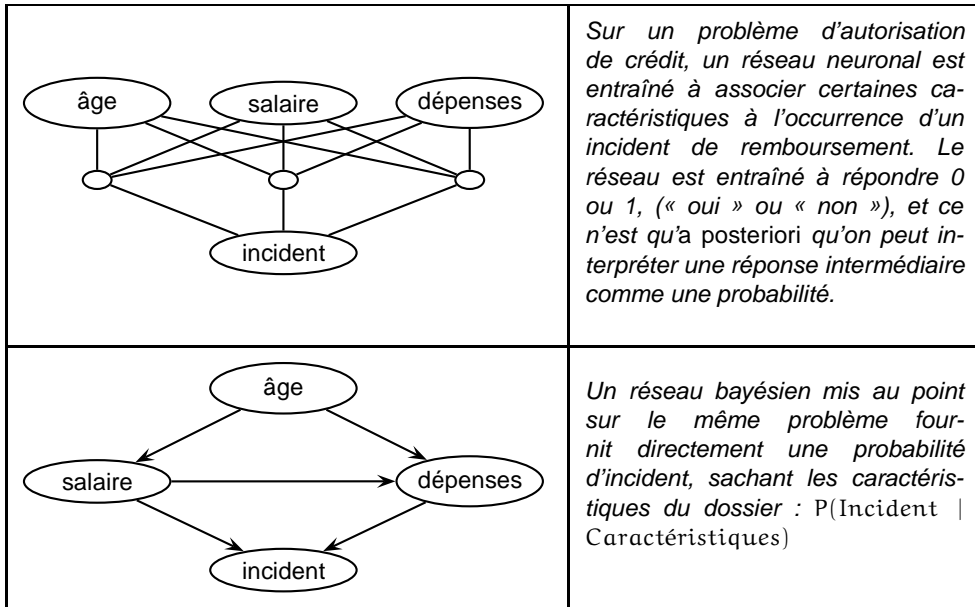


FIG. 7.2 Scoring et probabilités

7.1.2 Représentation des connaissances

► Un formalisme unificateur

La plupart des applications qui relèvent des réseaux bayésiens sont des applications d'aide à la décision. Par nature, ces applications intègrent un certain degré d'incertitude, qui est très bien pris en compte par le formalisme probabiliste des réseaux bayésiens.

Par exemple, les applications de data mining sont en général construites sur le schéma suivant. On utilise une base de données pour mettre au point un modèle prédictif. Par définition, une prévision comporte une part d'incertitude. Or la décision, elle, doit souvent être binaire : dans une application de scoring, on doit par exemple accorder ou refuser le crédit. La façon la plus naturelle d'interpréter un score est donc une probabilité (dans l'exemple du scoring, une probabilité de défaillance).

Les techniques disponibles pour traiter ce genre de problème (modèles de régression, réseaux de neurones, arbres de décision) ne sont pas construites sur un formalisme de probabilités. C'est *a posteriori* qu'on attribue en général une interprétation en termes de probabilités de la prévision d'un réseau neuronal ou d'un arbre de décision.

Les réseaux bayésiens ne sont qu'une représentation d'une distribution de probabilités. C'est une telle distribution que l'on représente à partir de connaissances explicites ou qu'on approche à partir d'une base de données, et c'est à partir de la distribution approchée que l'on effectue des inférences. Toute prévision issue d'un réseau bayésien est donc par construction une probabilité.

De plus, les réseaux bayésiens permettent de considérer dans un même formalisme la représentation de modèles de causalités et les statistiques multivariées. Il en est de même des techniques les plus utilisées pour le *data mining* comme les arbres de décision ou les réseaux de neurones, qui peuvent également être représentés au sein de ce formalisme.

► Une représentation des connaissances lisible

Les deux propriétés fondamentales des réseaux bayésiens sont, d'abord, d'être des graphes orientés, c'est-à-dire de représenter des causalités et non des simples corrélations, et, ensuite, de garantir une correspondance entre la distribution de probabilité sous-jacente et le graphe associé.

D'après le théorème d'indépendance graphique, que nous avons démontré dans la partie précédente, les relations de causalité et d'indépendance qui peuvent être lues sur le graphe sont également vraies dans la distribution sous-jacente.

Considérons le cas d'une application de *data mining*, où l'on cherche à comprendre les interrelations entre des variables contenues dans une base de données de clients, par exemple. Si l'on se trouve dans le cas où le réseau est entièrement mis au point à partir des données (cas de l'apprentissage de la structure et des paramètres), cela signifie que l'on va disposer d'une visualisation graphique de ces interrelations. Avant même d'utiliser ce réseau pour effectuer des inférences, on va disposer d'une visualisation de la connaissance, directement lisible et interprétable par des experts du domaine.

7.1.3 Utilisation de connaissances

► Une gamme de requêtes très complète

L'utilisation première d'un réseau bayésien est le calcul de la probabilité d'une hypothèse connaissant certaines observations. C'est sur cette requête élémentaire que nous avons abordé les calculs dans le chapitre 1. Cependant, les possibilités offertes par les algorithmes d'inférence permettent

d'envisager une gamme de requêtes très complète, qui peut être extrêmement intéressante dans certains types d'applications.

Tout d'abord, il n'y a aucune réelle contrainte sur les informations nécessaires pour être en mesure de calculer la probabilité d'un fait : on peut connaître exactement la valeur d'une variable, savoir qu'elle est égale à l'une ou l'autre de deux valeurs, ou encore savoir avec certitude qu'une de ses valeurs possibles est exclue. Dans tous les cas, l'inférence est possible, et la nouvelle information permet de raffiner les conclusions.

Il n'y a pas d'entrées ni de sorties dans un réseau bayésien (ou de variables indépendantes et dépendantes). Le réseau peut donc être utilisé pour déterminer la valeur la plus probable d'un nœud en fonction d'informations données (prévoir, ou sens entrées vers sorties), mais également pour connaître la cause la plus probable d'une information donnée (expliquer, ou sens sorties vers entrées). En termes d'inférences, cette dernière requête s'appelle explication la plus probable et revient, l'état de certaines variables étant observé, à rechercher l'état des autres variables pour lequel ce qui a été observé était le plus probable. Parmi les autres requêtes importantes, l'analyse de sensibilité à une information mesure comment la probabilité d'une hypothèse s'accroît quand on a fait une observation. Certaines observations peuvent ainsi être considérées comme inutiles, suffisantes, ou cruciales, par rapport à une hypothèse donnée.

Le mécanisme de propagation peut être également utilisé pour déterminer l'action la plus appropriée à effectuer, ou l'information la plus pertinente à rechercher. Considérons par exemple un problème de diagnostic, dans lequel manquent plusieurs des données qui permettraient de conclure. Le mécanisme de propagation dans un réseau bayésien permet de connaître la donnée dont la connaissance apporterait le maximum d'informations. Dans le cas où la recherche de chaque donnée a un coût, il est possible de rechercher la solution optimale en tenant compte de ce coût. De plus, il est possible de chercher également une séquence optimale d'actions ou de requêtes.

► Optimisation d'une fonction d'utilité

Imaginons un problème de classification, par exemple un problème de détection de fraudes sur des cartes bancaires, ou dans l'utilisation de services de télécommunications. Rechercher le système qui donne, avec la meilleure fiabilité possible, la probabilité de fraude n'est peut-être pas l'objectif réel de ce type d'application. En effet, ce qu'on cherche ici à optimiser est une utilité économique. Sachant que les fausses alarmes aussi bien que les fraudes manquées ont un coût, l'objectif est bien de minimiser le coût global. Une version spécifique des réseaux bayésiens, appelée diagramme

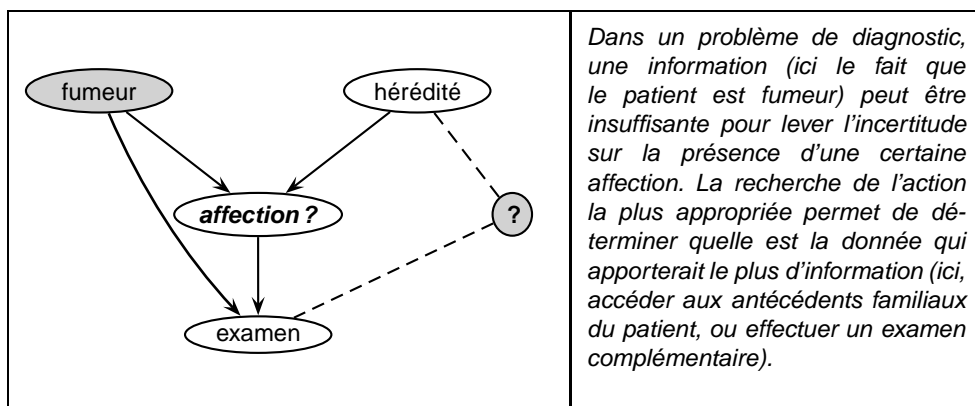


FIG. 7.3 Requête élaborée dans un réseau bayésien

d'influences, permet de les adapter à ce type de problème. Dans les diagrammes d'influence, on ajoute aux nœuds qui représentent des variables, deux autres types de nœuds :

- les nœuds de décision, figurés par des carrés ;
- un nœud d'utilité, figuré par un losange.

Le graphe ci-dessus représente un diagramme d'influence pour un problème de détection de fraude sur une carte bancaire. Les variables représentées sont les suivantes :

- La variable F est binaire et représente le fait qu'il y a ou non fraude.
- La variable B représente le résultat d'une vérification effectuée sur une base de données. Cette variable a trois modalités : le contrôle est négatif, positif, ou non effectué.
- La variable P a également trois modalités, et représente le résultat d'un contrôle d'identité du porteur.
- Le nœud de décision D représente la décision d'effectuer les contrôles complémentaires B et P. Ce nœud a donc également trois modalités : n'effectuer aucun test, effectuer le test B, ou effectuer les deux tests B et P.
- Le nœud de décision A représente la décision d'autoriser la transaction, et est donc binaire.
- Le nœud d'utilité V est une fonction de l'ensemble des variables précédentes, représentant le coût de la situation.

En outre, on suppose connus le montant de la transaction et le coût de chaque contrôle, et les tables de probabilités conditionnelles reliant les variables entre elles. L'objectif est de prendre les bonnes décisions D et A ; autrement dit, de prendre les décisions qui minimisent l'espérance mathématique de V.

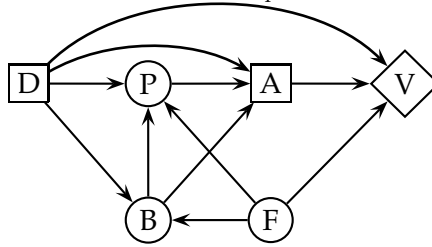


FIG. 7.4 Un diagramme d'influence pour la fraude sur carte bancaire

7.1.4 Limites des réseaux bayésiens

► Un recul encore insuffisant pour l'apprentissage

Dans la mesure où elle s'est surtout développée dans le cadre des systèmes experts, la technique des réseaux bayésiens n'a pas immédiatement intégré l'ensemble de la problématique de l'apprentissage, comme cela avait été le cas des réseaux neuronaux. Aujourd'hui, l'essentiel de la littérature sur l'apprentissage avec des réseaux bayésiens ignore le problème de la capacité de généralisation d'un modèle, et des précautions que cela implique au moment de la construction du modèle. La prise en compte de ce problème peut s'effectuer par le choix du critère de recherche ou de distance des distributions de probabilité. En effet, comme on l'a vu dans la partie précédente, l'apprentissage de réseaux bayésiens revient à rechercher parmi un ensemble de distributions, celle la plus proche possible, en un certain sens, de la distribution représentée par les données. En limitant l'ensemble de recherche, on peut éviter le problème de surapprentissage, qui revient dans ce cas à calquer exactement la distribution représentée par les exemples.

► Utilisation des probabilités

L'utilisation des graphes de causalités est, on l'a dit, une approche très intuitive. Nous avons montré que l'utilisation des probabilités pour rendre ces modèles quantitatifs était justifiée. Il reste cependant que la notion de probabilité, est, au contraire, assez peu intuitive. Il est en effet assez facile de construire des paradoxes fondés sur des raisonnements probabilistes. Les modèles déterministes, formulés en termes d'entrées et de sorties, comme les modèles de régression, les réseaux de neurones, ou les arbres de décision, même s'ils peuvent être réinterprétés dans le cadre d'un formalisme probabiliste, restent d'un abord plus facile.

► **Lisibilité des graphes**

En effet, même si la connaissance manipulée dans les réseaux bayésiens, ou extraites des données par les algorithmes d'apprentissage associés est lisible puisque représentée sous forme de graphes, elle reste moins lisible que celle représentée par un arbre de décision, par exemple, surtout si ce graphe présente un grand nombre de nœuds. Notons aussi que l'information représentée par le graphe est la structure des causalités. Les probabilités ne sont pas représentables, et on n'a donc pas idée, à la simple lecture du graphe, de l'importance d'un arc donné. La figure 9.4 page 239 dans l'étude de cas sur le risque global d'une entreprise (GLORIA) donne un aperçu d'un graphe complexe.

► **Les variables continues**

L'essentiel des algorithmes développés pour l'inférence et l'apprentissage dans les réseaux bayésiens, aussi bien que les outils disponibles sur le marché pour mettre en œuvre ces algorithmes utilisent des variables discrètes. En effet, comme nous l'avons vu dans la partie technique, la machinerie des algorithmes d'inférence est essentiellement fondée sur une algèbre de tables de probabilités. De même, les algorithmes d'apprentissage modélisent en général les distributions de probabilité des paramètres contenus dans les tables du réseau, c'est-à-dire de probabilités discrètes. Même s'il est théoriquement possible de généraliser les techniques développées aux variables continues, il semble que la communauté de recherche travaillant sur les réseaux bayésiens n'a pas encore vraiment intégré ces problèmes. Cela pénalise cette technologie, en particulier pour des applications de data mining où variables continues et discrètes cohabitent.

► **La complexité des algorithmes**

La généralité du formalisme des réseaux bayésiens aussi bien en termes de représentation que d'utilisation les rend difficiles à manipuler à partir d'une certaine taille. La complexité des réseaux bayésiens ne se traduit pas seulement en termes de compréhension par les utilisateurs. Les problèmes sous-jacents sont pratiquement tous de complexité non polynomiale, et conduisent à développer des algorithmes approchés, dont le comportement n'est pas garanti pour des problèmes de grande taille.

Connaissances	Analyse de données	Réseaux neuronaux	Arbres de décision	Systèmes experts	Réseaux bayésiens
ACQUISITION					
Expertise seulement				★	
Données seulement	+	★	+		+
Mixte	+	+	+		★
Incrémental		+			★
Généralisation	+	★	+		+
Données incomplètes		+			★
REPRÉSENTATION					
Incertitude				+	★
Lisibilité	+		+	+	★
Facilité		+	★		
Homogénéité					★
UTILISATION					
Requêtes élaborées	+			+	★
Utilité économique	+	+			★
Performances	+	★			

TAB. 7.1 *Avantages comparatifs des réseaux bayésiens*

7.1.5 Comparaison avec d'autres techniques

Du point de vue des applications, les avantages et inconvénients des réseaux bayésiens par rapport à quelques-unes des techniques concurrentes peuvent se résumer sur le tableau ci-dessus. Nous avons regroupé avantages et inconvénients selon les trois rubriques utilisées précédemment, l'acquisition, la représentation et l'utilisation des connaissances. La représentation adoptée est la suivante :

- À chaque ligne correspond une caractéristique, qui peut être un avantage, ou la prise en compte d'un problème spécifique.
- Si la technique considérée permet de prendre en compte ce problème, ou présente cet avantage, un signe + est placé dans la case correspondante.
- Un signe ★ est placé dans la case de la meilleure technique du point de vue de la caractéristique considérée.

7.2 Où utiliser des réseaux bayésiens ?

Les propriétés étudiées ci-dessus nous permettent de définir les caractéristiques générales d'une application où il est intéressant d'utiliser des réseaux bayésiens en les préférant à une autre technique. Les types d'ap-

plications relevant de cette approche sont listés plus loin.

7.2.1 Caractéristiques générales

► Une connaissance explicite ou implicite du domaine

Dans la mesure où un réseau bayésien peut être construit soit à partir de données, par apprentissage, soit à partir d'une modélisation explicite du domaine, il suffit que l'une ou l'autre des formes de connaissances ou une combinaison des deux soit disponible pour pouvoir envisager d'utiliser cette technique dans une application.

En partant d'une connaissance explicite même incomplète, et en utilisant la capacité d'apprentissage incrémental des réseaux bayésiens, on peut développer une approche de modélisation en ligne, c'est-à-dire sans archiver les exemples mêmes. En effet, un réseau bayésien n'est rien d'autre que la représentation d'une distribution de probabilité. Si la structure de cette distribution est imposée, on peut directement calculer l'impact de chaque nouvel exemple sur les paramètres de cette distribution.

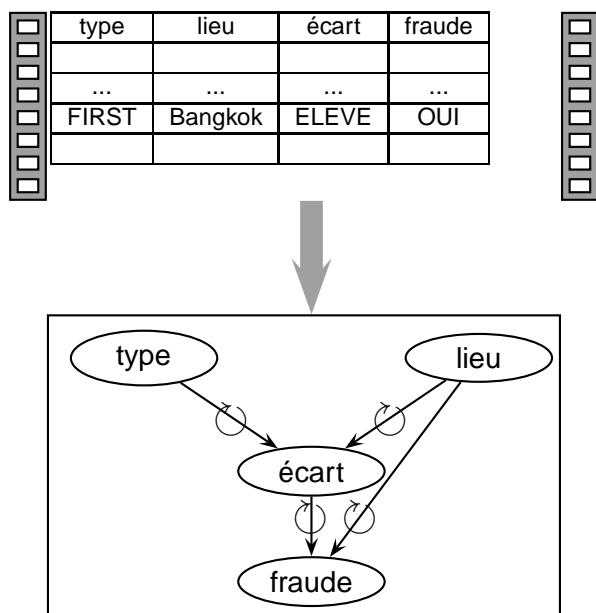


FIG. 7.5 Modélisation en ligne pour la détection de fraudes

Ce type d'approche peut être envisagé dans des applications de prévision de comportements d'achat ou de détection de fraudes, par exemple dans le cas du commerce électronique. Le schéma de la figure 7.5 page précédente montre un exemple très simple d'une telle application :

- On cherche à modéliser la distribution de probabilité liant le type d'une carte de crédit, le lieu de la transaction, l'écart du montant de la transaction par rapport à la moyenne et l'existence d'une fraude.
- La structure de cette distribution est supposée fixée.
- Le flux des transactions enregistrées modifie les tables de probabilités conditionnelles.

Cet exemple suppose cependant une historisation partielle des informations, puisque le retour sur la fraude effective ne peut intervenir immédiatement.

► Une utilisation complexe ou évolutive

En termes d'utilisation du modèle, l'avantage essentiel des réseaux bayésiens par rapport aux autres techniques est de permettre une formalisation complète d'un domaine de connaissances sous forme de graphe causal. Ce graphe peut être utilisé ensuite pour effectuer des raisonnements, en formulant des requêtes relativement complexes.

Cependant, cette technologie reste aujourd'hui relativement difficile à maîtriser pour des problèmes de grande taille. Il nous semble donc qu'elle ne doit être considérée que là où les techniques plus simples ne peuvent pas donner satisfaction.

Par exemple, dans un problème de prévision ou de classification spécifié de façon claire, et pour lequel la connaissance des règles sous-jacentes n'est pas essentielle, il nous semble préférable d'utiliser un modèle de régression ou un réseau de neurones.

En revanche, pour une application de data mining, au sens premier du terme, c'est-à-dire au sens où l'on recherche des relations *a priori* non connues entre des données, les méthodes d'apprentissage dans les réseaux bayésiens constituent selon nous une approche très prometteuse.

Enfin, dans certaines applications, la formulation initiale du besoin peut masquer des évolutions pour lesquelles des requêtes complexes sur le modèle peuvent s'avérer nécessaires. Considérons par exemple une application de *credit scoring*. Initialement formulée comme une application simple de classification, elle est mise en œuvre sous forme de réseau de neurones. Après quelques mois d'utilisation, on s'aperçoit d'une augmentation significative du taux de refus d'autorisation. L'interprétation de ce problème peut se révéler difficile sans un modèle capable d'explications.

7.2.2 Classification des applications par types

Un réseau bayésien est un moyen de représenter la connaissance d'un système. Une telle représentation n'est bien entendu pas une fin en soi ; elle s'effectue, selon les contextes, dans le but de :

- **prévoir** le comportement du système ;
- **diagnostiquer** les causes d'un phénomène observé dans le système ;
- **contrôler** le comportement du système ;
- **simuler** le comportement du système ;
- **analyser des données** relatives au système ;
- prendre des **décisions** concernant le système.

Ces différents types d'applications reposent en général sur deux types de modèles : les modèles symboliques pour le diagnostic, la planification, et les modèles numériques pour la classification, la prévision, le contrôle.

Comme nous l'avons vu ci-dessus, les réseaux bayésiens autorisent les deux types de représentation et d'utilisation des connaissances. Leur champ d'application est donc vaste, d'autant que le terme système s'entend ici dans son sens le plus large. Il peut s'agir, pour donner quelques exemples, du contenu du chariot d'un client de supermarché, d'un navire de la Marine, du patient d'une consultation médicale, du moteur d'une automobile, d'un réseau électrique ou de l'utilisateur d'un logiciel. Ajoutons que la communauté de chercheurs qui développent la théorie et les applications des réseaux bayésiens rassemble plusieurs disciplines scientifiques : l'intelligence artificielle, les probabilités et statistiques, la théorie de la décision, l'informatique et aussi les sciences cognitives. Ce facteur contribue à la diffusion et donc à la multiplicité des applications des réseaux bayésiens.

► Modèles symboliques

Par rapport aux systèmes à base de règles déterministes, le plus souvent utilisés dans les systèmes experts, les réseaux bayésiens permettent d'intégrer l'incertitude dans le raisonnement.

Ils sont donc adaptés aux problèmes où l'incertitude est présente, que ce soit dans les observations, ou dans les règles de décision.

Les systèmes de diagnostic sont ceux qui utilisent le plus complètement les possibilités des réseaux bayésiens, en particulier en ce qui concerne les capacités d'explication, de simulation, etc. Un avantage spécifique des réseaux bayésiens dans les problèmes de diagnostic est de pouvoir détecter plusieurs pannes simultanées. Les techniques déterministes comme les arbres de décision conduisent le plus souvent à un seul diagnostic à la fois.

Certaines applications de planification peuvent également utiliser des

réseaux bayésiens, mais utilisés en quelque sorte comme sous-systèmes, permettant de déterminer les actions dont la faisabilité ou le succès ont une bonne probabilité.

Les réseaux bayésiens sont en revanche moins adaptés aux applications apparentées à la résolution de problèmes ou à la démonstration de théorèmes.

► Modèles numériques

Les systèmes de classification mettent en général en œuvre des architectures simplifiées de réseaux (arbres ou polyarbres). Des études et des applications récentes montrent que les systèmes de classification basés sur des arbres bayésiens donnent des résultats en général significativement meilleurs que les algorithmes de classification de type arbre de décision (C4.5, C5).

Les méthodes d'apprentissage de structure dans les réseaux bayésiens permettront de donner tout son sens au terme de data mining. S'il s'agit en effet de rechercher des relations entre des variables sans *a priori*, ni les réseaux de neurones, ni les arbres de décision ne sont adaptés à ce type de problème. Dans les applications de modélisation numérique, comme la prévision, le contrôle ou l'estimation, il nous semble en revanche que le formalisme global des réseaux bayésiens est trop lourd pour être utilisé tel quel, du moins dans un premier temps. Cela ne signifie pas pour autant qu'il ne puisse pas contribuer à de telles applications. Ainsi, une étude récente a permis d'améliorer significativement la performance de réseaux neuronaux en prévision, en optimisant le choix des paramètres d'apprentissage grâce à un réseau bayésien simple. À terme, l'unification des algorithmes d'apprentissage permettra sans doute d'intégrer un modèle neuronal de prévision à l'intérieur d'un système plus global, où pourront être optimisés simultanément, soit des paramètres d'apprentissage, soit des paramètres de la décision basée sur la prévision.

7.2.3 Classification des applications par domaines

► Santé

Les premières applications des réseaux bayésiens ont été développées dans le domaine du diagnostic médical.

Les réseaux bayésiens sont particulièrement adaptés à ce domaine parce qu'ils offrent la possibilité d'intégrer des sources de connaissances hétérogènes (expertise humaine et données statistiques), et surtout parce que leur

capacité à traiter des requêtes complexes (explication la plus probable, action la plus appropriée) peuvent constituer une aide véritable et interactive pour le praticien.

Le système Pathfinder, développé au début des années 1990 a été conçu pour fournir une assistance au diagnostic histopathologique, c'est-à-dire basé sur l'analyse des biopsies. Il est aujourd'hui intégré au produit Intelipath, qui couvre un domaine d'une trentaine de types de pathologies. Ce produit est commercialisé par l'éditeur américain Chapman et Hall, et a été approuvé par l'*American Medical Association*.

Dans le domaine de la santé, une application intéressante des algorithmes issus des réseaux bayésiens a permis d'améliorer considérablement la recherche de la localisation de certains gènes, dans le cadre du projet Human Genome. Nous reviendrons sur cette application dans la section suivante.

► Industrie

Dans le domaine industriel, les réseaux bayésiens présentent également certains avantages par rapport aux autres techniques d'intelligence artificielle. Leur capacité réelle d'apprentissage incrémental, c'est-à-dire d'adaptation de la connaissance en fonction des situations rencontrées, en fait les contrôleurs idéaux de systèmes autonomes ou de robots adaptatifs.

En effet, la propriété essentielle d'un système autonome, pour pouvoir « survivre », est de s'adapter aux modifications structurelles de son environnement. La capacité du système à gérer ses propres altérations, en particulier la perte de certaines fonctions, est également importante. Ainsi, dans la situation où certains de ses capteurs ou effecteurs sont endommagés, le système doit être capable de mettre à jour son domaine de viabilité, c'est-à-dire de réévaluer les capacités d'action qu'il lui reste, malgré le dommage qu'il a subi.

C'est cette idée qui a été mise en œuvre par la société danoise Hugin, considérée comme l'un des pionniers dans le développement des réseaux bayésiens. Hugin a développé pour le compte de Lockheed Martin le système de contrôle d'un véhicule sous-marin autonome. Ce système évalue en permanence les capacités du véhicule à réagir à certains types d'événements. De cette façon, en fonction des capacités qui sont cruciales pour le reste de la mission, le système peut prendre des décisions qui vont de la simple collecte d'informations complémentaires, à la modification de la mission, ou jusqu'à l'abandon de celle-ci.

Transposant cette idée de contrôle de systèmes autonomes du monde réel à l'univers virtuel des systèmes et réseaux informatiques, les réseaux

bayésiens devraient également équiper les agents intelligents. Comme nous l'avons déjà mentionné, le diagnostic est un des autres domaines de prédilection des réseaux bayésiens dans l'industrie, en particulier grâce à l'utilisation des requêtes avancées sur les réseaux. Ce domaine est aujourd'hui l'un des plus développés en termes d'applications opérationnelles des réseaux bayésiens (Hewlett-Packard, General Electric, Ricoh, etc.)

► Défense

Comme pour beaucoup de techniques issues de l'intelligence artificielle, c'est grâce à la défense américaine que les réseaux bayésiens ont pu connaître leurs premiers développements.

La fusion de données est en particulier un domaine d'application privilégié des réseaux bayésiens, grâce à leur capacité à prendre en compte des données incomplètes ou incertaines, et à guider la recherche ou la vérification de ces informations.

La fusion de données peut se définir comme le processus qui consiste à inférer une information à laquelle on n'a pas directement accès, mais qui est relayée par une ou plusieurs sources imparfaites. Finalement, un détective privé qui affine ses conclusions à mesure que les indices se complètent est un spécialiste de la fusion de données.

Il est clair que cette approche est essentielle dans le domaine du renseignement, tactique ou stratégique. Par exemple, l'identification d'un navire ennemi est impossible directement. On va combiner des informations issues de systèmes de mesure, éventuellement brouillées, avec d'autres types de renseignements, également incertains. Les informations disponibles se complètent au fur et à mesure des efforts accomplis pour identifier ce navire, permettant de renforcer ou, au contraire, de réviser les conclusions effectuées.

Un exemple d'application dans la défense tactique est donné dans la section suivante.

► Banque/finance

Les applications dans le domaine de la banque et de la finance sont encore rares, ou du moins ne sont pas publiées. Mais cette technologie présente un potentiel très important pour un certain nombre d'applications relevant de ce domaine, comme l'analyse financière, le scoring, l'évaluation du risque ou la détection de fraudes.

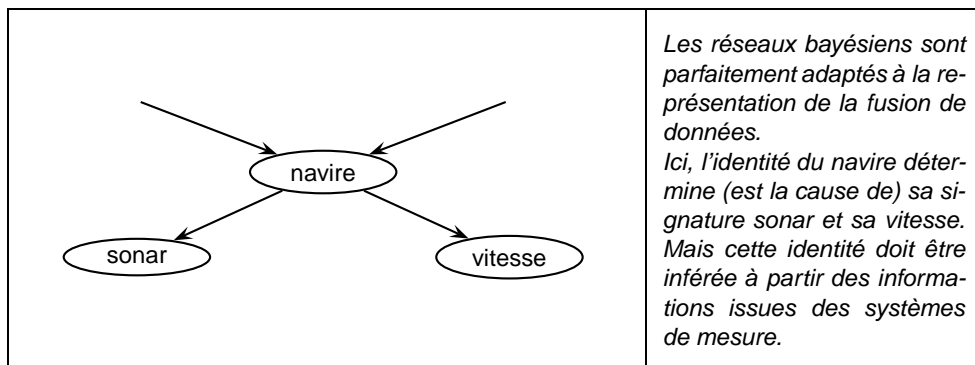


FIG. 7.6 Principe de la fusion de données par réseau bayésien

En premier lieu, les réseaux bayésiens offrent un formalisme unifié pour la manipulation de l'incertitude, autrement dit du risque, dont la prise en compte est essentielle dès qu'il s'agit de décision financière.

Ensuite, la possibilité de coupler expertise et apprentissage est ici très importante, non seulement parce que les deux sources de connaissances sont en général disponibles dans ce domaine, mais aussi et surtout parce que cette capacité peut aider à répondre au problème des changements structurels d'environnement.

Traitées dans les années 1980 avec des systèmes experts, des applications comme l'analyse financière, le *scoring* ou la détection de fraudes ont été progressivement considérées comme relevant du domaine de la modélisation quantitative, et donc abordées par des techniques comme les réseaux neuronaux ou les arbres de décision, techniques quantitatives qui se révèlent incapables de prendre en compte par elles-mêmes la révision des modèles.

L'exemple de l'autorisation des transactions sur cartes bancaires est assez significatif. L'un des premiers systèmes experts développés dans ce domaine fut l'*Authorizer Assistant* d'*American Express*, au début des années 80. Dès la fin de la décennie, la société californienne HNC (*Hecht-Nielsen Neurocomputing*) devient le leader des systèmes de détection de fraudes sur cartes bancaires. Son système *Falcon* équipe la plupart des émetteurs de cartes aux États-Unis. Fondé initialement sur une technologie de réseaux neuronaux, le système *Falcon* a récemment évolué pour y intégrer... un système expert ! Pourquoi ? La fraude est, presque par définition, un phénomène évolutif, qui s'adapte aux parades qui y sont opposées. Un modèle construit à partir de données historiques a donc nécessairement une durée de vie limitée dans un tel environnement. Donc, même si les réseaux de neurones étaient

la technique la plus fiable pour identifier les comportements frauduleux, comme ils reposent sur le traitement de données historiques, ils ne peuvent s'adapter assez vite aux changements de ces comportements.

On retrouve la même problématique dans la finance de marchés, où les modèles de prévision ou de gestion mis au point sur des données historiques ne peuvent s'adapter aux changements structurels brusques des marchés.

Récemment, les nouveaux accords de Bâle II ont ouvert un nouveau champ d'application très significatif pour les réseaux bayésiens dans le domaine bancaire. Ces accords fixent les nouvelles règles que doivent appliquer les banques pour la détermination de leurs exigences en fonds propres. Ces fonds propres doivent être dimensionnés de façon à couvrir à un niveau de probabilité élevé les différents types de risques encourus par la banque : risques de crédit, risques de marché et risques opérationnels.

Le risque opérationnel a été défini par l'accord de Bâle II de façon générale comme « le risque de pertes provenant de processus internes inadéquats ou défaillants, de personnes et systèmes ou d'événements externes », et de façon spécifique en identifiant sept thèmes principaux de risque, comme la fraude, la relation avec les clients ou le personnel, les systèmes d'information, etc. La prise en compte de ces risques est en général très difficile, car les plus significatifs concernent des événements rares mais de fort impact. Comme de nombreux spécialistes de la gestion du risque l'ont mis en avant, en particulier [Ale02], l'utilisation de modèles bayésiens est particulièrement adaptée pour plusieurs raisons :

- Les réseaux bayésiens permettent de coupler les connaissances des experts et les données disponibles.
- Ils permettent de conditionner les risques et donc de mieux évaluer les pertes encourues.
- Ils permettent d'identifier des leviers de réduction de risque.
- Les modèles établis sont transparents et facilement auditables par les organismes de contrôle.

Gageons que l'utilisation des réseaux bayésiens deviendra probablement l'une des méthodes de référence pour la modélisation du risque opérationnel.

L'étude de cas sur la modélisation du risque global d'une entreprise (méthode GLORIA) que nous présentons dans le chapitre 9 permet d'avoir un aperçu de la démarche qui peut être adoptée pour une telle application, même si les objectifs visés sont plus qualitatifs.

► Marketing

Ce que l'on appelle aujourd'hui le data mining, est probablement le domaine où le potentiel des réseaux bayésiens est le plus élevé. Le data mining est défini par certains comme l'extraction automatique à partir de bases de données d'informations *a priori* inconnues et à valeur prédictive. Nous préférons le définir comme l'utilisation rationnelle de l'information contenue dans les données pour la prise de décision.

Quelle que soit la définition retenue, il reste que le développement actuel du data mining s'explique essentiellement par les applications dans le domaine du marketing, et que les réseaux bayésiens sont parfaitement adaptés à ces applications.

Le marketing est en train d'évoluer vers une gestion de plus en plus fine et individualisée du capital client, considéré comme un nouvel actif de l'entreprise. Les applications de prévision, de fidélisation, d'analyse du risque, d'anticipation des besoins, de ciblage d'actions s'inscrivent toutes dans cette démarche.

Toutes les caractéristiques des réseaux bayésiens sont autant d'atouts pour ces types d'applications :

- La gestion de l'incertitude, car évidemment toutes les actions marketing sont prises dans un contexte d'incertitude, où l'on recherche avant tout à augmenter la probabilité de succès.
- La capacité à intégrer des données incomplètes au cours de l'apprentissage, car les données utilisées dans le data mining appliqué au marketing proviennent souvent de sources déclaratives, de qualité approximative.
- L'apprentissage incrémental, car les relations évoluent avec le temps.
- La gestion de requêtes complexes, comme l'analyse de sensibilités, la recherche de l'action la plus appropriée.

Notons de plus que l'utilisation des réseaux bayésiens permet également d'envisager des applications de data mining pour de petites bases de données. Ce problème, qui reste rarement considéré aujourd'hui, peut cependant être très réel dans certaines applications. Comment faire pour tirer parti le plus rapidement possible de la connaissance à partir des premiers cas disponibles ? L'intégration avec de la connaissance *a priori* peut être une réponse à ce problème.

Nous pensons que le data mining, et en particulier ses applications dans le domaine du marketing seront l'un des moteurs principaux du développement des réseaux bayésiens dans un futur proche. Certains indicateurs, que nous analysons dans la section suivante sur l'offre commerciale et la recherche, semblent confirmer cette prévision.

► Informatique

Nous avons vu ci-dessus que certaines des caractéristiques des réseaux bayésiens en faisaient des systèmes de contrôle idéaux pour des systèmes autonomes dans des environnements changeants. Ces propriétés sont également valables pour équiper les agents logiciels, locaux à une machine, ou autonomes sur des réseaux ou sur Internet. Un agent logiciel est une application qui réalise de façon autonome une mission qui lui a été assignée par un utilisateur, ou par un autre agent. Les caractéristiques principales des agents intelligents sont donc :

- **L'autonomie.** Cela implique en particulier que l'agent doit pouvoir accomplir sa tâche sans se reporter systématiquement à son donneur d'ordre, et ce, même si des événements imprévus surviennent.
- **La motivation.** Un agent est dirigé par un but et doit éventuellement définir son propre plan d'action pour atteindre son but.
- **La réactivité.** Un agent doit pouvoir modifier son comportement lorsqu'une nouvelle information devient disponible.
- **L'adaptativité.** Un agent doit être capable d'intégrer les modifications de son environnement. Ceci est particulièrement vrai pour les agents opérant sur Internet ou d'autres réseaux, dont l'environnement est par nature instable.

Les réseaux bayésiens sont probablement l'une des technologies les plus adaptées pour construire l'intelligence des agents. Ils assurent en effet les différentes propriétés présentées précédemment :

- L'autonomie est représentée par la capacité des réseaux bayésiens de fournir des décisions en présence d'incertitude, ou en l'absence de certaines informations.
- La motivation peut être représentée par certains types d'inférences, ou par un système de planification.
- La réactivité est le principe même de l'inférence dans les réseaux bayésiens (révision de la conclusion).
- L'adaptation à l'environnement est rendue possible par les capacités d'apprentissage incrémental des réseaux bayésiens.

La compacité de la représentation de la connaissance autorisée par les réseaux bayésiens est aussi un avantage pour en faire une intelligence embarquée.

L'utilisation de réseaux bayésiens dans les agents bureautiques a été largement développée par Microsoft dans les outils d'aide et de diagnostic pour son système d'exploitation Windows, à partir de Windows 98. De même, l'agent Office Assistant est un système d'aide proactif intégré dans Office, à partir de la version 97. Plusieurs agents de support technique de Microsoft ont également été développés dans le cadre du projet LUMIERE

du groupe DTAS (*Decision Theory and Adaptive Systems*).

L'application Vista, détaillée dans le chapitre suivant, peut également être considérée comme un agent intelligent, dont le rôle est de sélectionner les données présentées à un utilisateur en fonction de l'état du système physique qu'il doit superviser.

Les réseaux bayésiens constituent selon nous le modèle idéal pour embarquer de l'intelligence ou de la connaissance.

Embarquer de l'intelligence revient à doter un agent d'un équipement lui permettant de décider dans des environnements incertains, et de s'adapter lorsque ces environnements changent. Un module bayésien de prise de décision, éventuellement capable d'adaptation, est l'un des meilleurs équipements que l'on puisse fournir à un agent envoyé en mission sur Internet, ou sur d'autres types de réseau, où l'information est par nature incertaine et évolutive, voire manipulée.

► Gestion des connaissances

Dans la première partie de ce livre, nous avons montré comment les réseaux bayésiens pouvaient être construits simplement en cherchant à quantifier la représentation de graphes de causalités. Cette représentation graphique des domaines de connaissance reste la base des réseaux bayésiens.

Si les réseaux sont de taille raisonnable, cette représentation de la connaissance est très simple et intuitive, et permet d'envisager des échanges de modèles de connaissances sous forme de réseaux. Certaines expériences ont montré que l'utilisation de réseaux bayésiens permet de faciliter l'échange entre experts d'un domaine.

Le domaine de la gestion des connaissances, qui connaît un intérêt croissant, est donc également un champ d'application potentiel pour les réseaux bayésiens, dans la mesure où ceux-ci offrent un formalisme riche et intuitif de représentation de la connaissance.

7.3 Comment utiliser des réseaux bayésiens ?

La construction d'un réseau bayésien s'effectue en trois étapes essentielles, qui sont présentées sur la figure 7.7 ci-après.

Chacune des trois étapes peut impliquer un recueil d'expertise, au moyen de questionnaires écrits, d'entretiens individuels ou encore de séances de *brainstorming*. Préconiser, dans un cadre général, l'une ou l'autre de ces approches serait pour le moins hasardeux ; les chapitres suivants montre-

ront quels choix ont été retenus dans plusieurs utilisations réelles des réseaux bayésiens.

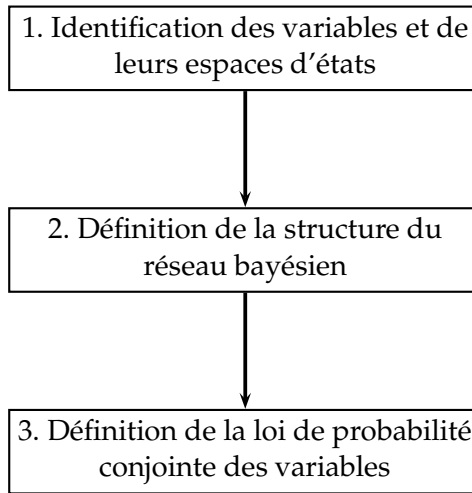


FIG. 7.7 Étapes de construction d'un réseau bayésien

7.3.1 Identification des variables et de leurs espaces d'états

La première étape de construction du réseau bayésien est la seule pour laquelle l'intervention humaine est absolument indispensable. Il s'agit de déterminer l'ensemble des variables X_i , catégorielles ou numériques, qui caractérisent le système. Comme dans tout travail de modélisation, un compromis entre la précision de la représentation et la maniabilité du modèle doit être trouvé, au moyen d'une discussion entre les experts et le modélisateur.

Lorsque les variables sont identifiées, il est ensuite nécessaire de préciser l'*espace d'états* de chaque variable X_i , c'est-à-dire l'ensemble de ses valeurs possibles.

La majorité des logiciels de réseaux bayésiens ne traite que des modèles à variables discrètes, ayant un nombre fini de valeurs possibles. Si tel est le cas, il est impératif de discrétiser les plages de variation des variables continues. Cette limitation est parfois gênante en pratique, car des discrétisations trop fines peuvent conduire à des tables de probabilités de grande taille, de nature à saturer la mémoire de l'ordinateur.

7.3.2 Définition de la structure du réseau bayésien

La deuxième étape consiste à identifier les liens entre variables, c'est-à-dire à répondre à la question : pour quels couples (i, j) la variable X_i influence-t-elle la variable X_j ?

Dans la plupart des applications, cette étape s'effectue par l'interrogation d'experts. Dans ce cas, des itérations sont souvent nécessaires pour aboutir à une description consensuelle des interactions entre les variables X_i . L'expérience montre cependant que la représentation graphique du réseau bayésien est dans cette étape un support de dialogue extrêmement précieux.

Un réseau bayésien ne doit pas comporter de circuit orienté ou boucle (figure 7.8). Cependant, le nombre et la complexité des dépendances identifiées par les experts laissent parfois supposer que la modélisation par un graphe sans circuit est impossible. Il est alors important de garder à l'esprit que, quelles que soient les dépendances stochastiques entre des variables aléatoires discrètes, il existe toujours une représentation par réseau bayésien de leur loi conjointe. Ce résultat théorique est fondamental et montre bien la puissance de modélisation des réseaux bayésiens.

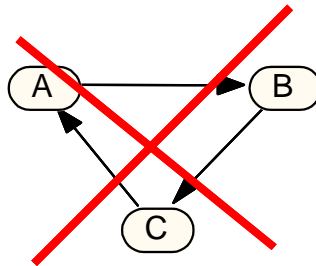


FIG. 7.8 Boucle dans un réseau bayésien

Lorsque l'on dispose d'une quantité suffisante de données de retour d'expérience concernant les variables X_i , la structure du réseau bayésien peut également être apprise automatiquement par le réseau bayésien, à condition bien sûr que le logiciel utilisé soit doté de la fonctionnalité adéquate.

7.3.3 Loi de probabilité conjointe des variables

La dernière étape de construction du réseau bayésien consiste à renseigner les tables de probabilités associées aux différentes variables.

Dans un premier temps, la connaissance des experts concernant les lois de probabilité des variables est intégrée au modèle.

Concrètement, deux cas se présentent selon la position d'une variable X_i dans le réseau bayésien :

- La variable X_i n'a pas de variable parente : les experts doivent préciser la loi de probabilité marginale de X_i .
- La variable X_i possède des variables parentes : les experts doivent exprimer la dépendance de X_i en fonction des variables parentes, soit au moyen de probabilités conditionnelles, soit par une équation déterministe (que le logiciel convertira ensuite en probabilités).

Le recueil de lois de probabilités auprès d'experts est une étape délicate du processus de construction du réseau bayésien. Typiquement, les experts se montrent réticents à chiffrer la plausibilité d'un événement qu'ils n'ont jamais observé.

Cependant, une discussion approfondie avec les experts, aboutissant parfois à une reformulation plus précise des variables, permet dans de nombreux cas l'obtention d'appréciations qualitatives. Ainsi, lorsqu'un événement est clairement défini, les experts sont généralement mieux à même d'exprimer si celui-ci est probable, peu probable, hautement improbable, etc. Il est alors possible d'utiliser une table de conversion d'appréciations qualitatives en probabilités, comme l'échelle de Lichtenstein et Newman proposé par [Ayy01, LP01]. La figure 7.9 ci-après représente graphiquement un extrait de cette table (les marges d'erreur associées à chaque probabilité sont figurées en gris foncé). Le développement des réseaux bayésiens a donné lieu à de nombreux travaux sur le thème de la correspondance entre les termes linguistiques et les probabilités quantitatives [RW99].

Le cas d'absence totale d'information concernant la loi de probabilité d'une variable X_i peut être rencontré. La solution pragmatique consiste alors à affecter à X_i une loi de probabilité arbitraire, par exemple une loi uniforme. Lorsque la construction du réseau bayésien est achevée, l'étude de la sensibilité du modèle à cette loi permet de décider ou non de consacrer davantage de moyens à l'étude de la variable X_i .

La quasi-totalité des logiciels commerciaux de réseaux bayésiens permet l'apprentissage automatique des tables de probabilités à partir de données. Par conséquent, dans un second temps, les éventuelles observations des X_i peuvent être incorporées au modèle, afin d'affiner les probabilités introduites par les experts.

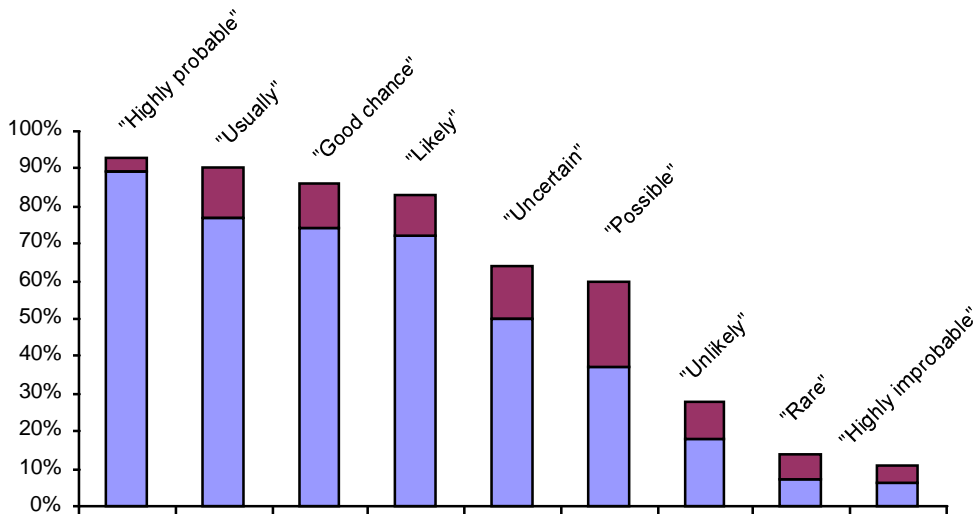


FIG. 7.9 Correspondance entre appréciations qualitatives et probabilités (échelle de Lichtenstein et Newman)

Il est rare en pratique que les données soient suffisamment nombreuses et fiables pour caractériser de manière satisfaisante la loi de probabilité conjointe des variables X_i . Cependant, si tel est le cas, l'apprentissage automatique des probabilités rend inutile la phase de renseignement du modèle par des probabilités expertes ; on peut alors se contenter, dans la phase initiale, d'attribuer à chaque variable une loi de probabilité uniforme.