# …

# Les réseaux Bayésiens et Les réseaux Bayésiens naïfs

## Motivation

Dans les problèmes de classification, un objet est représenté par un vecteur de caractéristiques ou attributs X, c’est-à-dire, X= [x1, x2, ..., xi, ..., xd]. En outre, chaque objet appartient à l'une des classes ou catégories possibles : { c1, c2, ..., ck }, par exemple dans notre cas les classes sont l’ensemble des chiffre {1 ,2,3,…,9}. En général, un système de classification suit deux étapes [1] : formation (apprentissage) et classification (test). Pendant la phase d'entraînement, le classifieur est entièrement conçu (mettre en place ses paramètres) en utilisant un ensemble d'échantillons d'apprentissage nommé ensemble d'entraînement (trainingset).

Dans la phase de test, le classifieur entraîné affecte l’objet d'entrée à l'une des classes en fonction de ses attributs. La précision du classificateur dépend à la fois du nombre et la pertinence d'échantillons d'apprentissage et les valeurs spécifiques de ces échantillons.

Nous considérons que les échantillons d'entraînement sont étiquetés (apprentissage supervisé), c'est-à-dire l'étiquette t d'un échantillon d'entraînement X représente la classe à laquelle appartient cet échantillon. La figure 2a montre la forme générale d’un dataset d’apprentissage dans la classification supervisée.

Bien évidemment, l'objectif de la conception d'un système de classification est de prédire les futurs échantillons d'essai (figure 2b) qui n'ont pas été forcément utilisés pendant la phase d'entraînement. Notre problème exige une reconnaissance des plaques d’immatriculation dans une séquence vidéo en temps réel où la prédiction est souvent coûteuse en termes de calcul (temps) et aussi en termes d’espace (mémoire) à cause de :

* Le flux de données en entrée à traiter est important.
* L’extraction de quelques attributs est couteuse.

Afin d’accélérer le processus de classification nous avons catégorisé les attributs en deux sous-ensembles :

* Le sous-ensemble d’attributs auxquels leur extraction est rapide.
* Le sous-ensemble d’attributs auxquels leur extraction est relativement lente.

Cette catégorisation permettra le classifieur dans un premier temps de faire la prédiction en se basant sur le premier sous-ensemble, l’exploitation des attributs du deuxième sous-ensemble via seulement quand l’objet n’est pas facile à prédire, c’est-à-dire le classifieur ne donne pas un degré de confiance suffisant. Cependant la plupart des méthodes de classification n’offrent pas cette flexibilité à moins qu’ils refassent l’entrainement.

Le classificateur bayésien naïf présente relativement une très bonne solution au problème de classification rapide et avec données incomplètes vu les avantages suivants [2], [3] :

1. Gère bien la classification en streaming c’est-à-dire quand on n’a pas forcements tous les attributs utilisés pendant la phase d’apprentissage pour pouvoir faire une prédiction, ce qui le plus important pour nous.
2. Très simple, rapide et facile à mettre en œuvre.
3. Si l'hypothèse d'indépendance conditionnelle du N.-B. est vérifiée, alors elle convergera plus rapidement que les modèles discriminants comme les réseaux de neurones artificiels ou les méthodes SVM.
4. Même si l'hypothèse du NB ne tient pas parfaitement, cela fonctionne très bien dans la pratique.
5. Besoin de moins de données d'entraînement vu qu’on dispose pas d’un dataset de taille importante.
6. Peut être utilisé à la fois pour les problèmes de classification binaire et multi-classes.
7. Peut faire des prédictions probabilistes ce qui est très significatif.
8. Non sensible aux attributs non pertinents.

Le réseau Bayésien Naïf est la forme la plus simple des réseaux Bayésiens, où tous les attributs sont indépendants en tenant compte la classe. C'est ce qu'on appelle l'indépendance conditionnelle. Dans les sections suivantes nous détaillerons la théorie liée aux réseaux Bayésiens en générale et en particulier le réseaux Bayésien naïf.

## Les réseaux Bayésiens (RBs)

### Introduction aux RBs

Les RBs représentent un outil puissant de décision, car ils sont capables de combiner l’aspect quantitatif sous forme de probabilités conditionnelles et l’aspect qualitatif par la spécification graphique des relations d’indépendance entre variables, ils s’appuient sur la théorie de graphes et celle du calcul probabiliste, ce qui explique l’appellation « modèles graphiques probabilistes » par certains auteurs pour désigner les RBs.

Les RBs s’appuient sur le théorème célèbre de Bayes. C’est un résultat de base en théorie des probabilités, issu des travaux de Thomas Bayes (1702-1761). Les résultats de ce théorème sont :

*(II.1)*

*(II.2)*

Le terme désigne la probabilité a priori de. Elle est « antérieure » au sens qu’elle précède toute information sur. est aussi appelée la probabilité marginale de. Le terme est appelé la probabilité a posteriori de sachant (ou encore de sachant). Elle est « postérieure », au sens qu’elle dépend directement de. Le terme, pour un connu, elle est appelée la fonction de vraisemblance de. De même, le terme est appelé la probabilité marginale ou a priori de.

La formalisation des RBs a commencé vers la fin des années 80 grâce aux travaux de Judea Pearl [4]. Depuis, l'intérêt pour les RBs dans la communauté de l'intelligence artificielle n'a pas cessé de grandir. L’idée clé consiste à explorer et à exploiter les relations d’indépendance statistique, plutôt que de s’intéresser à la structure causale qui ne peut être toujours dévoilée à partir de données. Les RBs sont notamment utilisés pour la représentation de connaissances incertaines ainsi que pour le raisonnement probabiliste à partir des données incomplètes.

Dans la suite de ce chapitre, nous visons en particulier à donner une vue d’ensemble sur les RBs où nous essayons de mettre l’accent sur la définition de la notion des RBs, les méthodes de construction des RBs, et leur application au raisonnement probabiliste.

### Préliminaires

#### Indépendance statistique

Considérons, un ensemble de variables aléatoires. Dans les modèles graphiques probabilistes, deux propriétés sont centrales : l'indépendance marginale et l'indépendance conditionnelle.

***Définition 1:*** *Deux variables aléatoires et sont dites (statistiquement) indépendantes si et seulement si on a :*

*(II.3)*

L'indépendance entre deux variables et, notée, implique qu’une information sur les valeurs prises par la variable n’a pas d’impact sur notre connaissance de la distribution de probabilité de la variable. Il découle de cette définition que est valide si et seulement si :

*(II.4)*

Une non-égalité implique que et sont statistiquement dépendantes. L’indépendance est une notion symétrique, si est indépendante de, alors est indépendante de

#### ****Indépendance statistique conditionnelle****

***Définition*2 :** *Deux variables et sont conditionnellement indépendantes sachant un sous-ensemble de variables, noté, si et seulement si:*

*(‎II.5)*

Le concept d’indépendance conditionnelle est symétrique, si est conditionnellement indépendante de sachant alors est également conditionnellement indépendante de sachant. Par rapport à la distribution de la conjointe et les distributions marginales, l’indépendance conditionnelle est vérifiée si on a :

*(‎II.6)*

Une non-égalité implique que et sont conditionnellement dépendantes sachant. Il est bien de mentionner que l’indépendance représente un cas spécial de l’indépendance conditionnelle, lorsque l’ensemble de conditionnement est vide.

### Calcul probabiliste

* ***Règle fondamentale***

*(‎II.7)*

Cette règle nous montre comment nous pouvons calculer la distribution conjointe des deux variables et lorsque nous disposons de la distribution et de la distribution conditionnelle.

* ***Règle de Bayes***

*(‎II.8)*

La règle de Bayes nous offre un moyen pour actualiser notre connaissance sur la distribution d’une variable sachant qu’on a une information sur la distribution d’une autre variable*.* Pour cette raison, est appelée la probabilité a priori de, alors que est appelée la probabilité a postériori de sachant ; la probabilité conditionnelle est appelée la vraisemblance de étant donné.

### Définition des RBs

***Définition 3 (Réseaux Bayésiens)***

Un Réseau Bayésien est un modèle graphique probabiliste défini par le couple où :

* est un graphe orienté et sans circuit (DAG pour Directed Acyclic Graph) dont les sommets correspondent à un ensemble de variables aléatoires. Les arcsreprésentent les relations d’indépendance conditionnelles existantes entre les variables.
* un ensemble de distributions de probabilités conditionnelles, une distribution de pour chaque instance de.

Les RBs permettent selon la structure du graphe une factorisation de la fonction de distribution de probabilité conjointe en un produit de distributions de probabilité conditionnelle de chaque nœud conditionnellement à ses parents dans le graphe :

*(‎II.9)*

Les travaux actuels liés aux RBs portent principalement sur les algorithmes d'inférence [12], [6], l'apprentissage de la structure du réseau [14], [15], l'apprentissage des paramètres du réseau [13], et les extensions des modèles des RBs. A l’heure actuelle, les algorithmes développés d’inférence exacte ou approchée et ceux d’apprentissage sont opérationnels pour le cas des RBs classiques de taille moyenne (réseaux statiques à variables discrètes).

***Exemple 1 d’un RB***

L’exemple affiché sur la Figure II.1 modélise les relations d’indépendances conditionnelles entres une variable de décision V (fait le voyage) et 3 autres variables modélisants les conditions climatiques à Alger qui sont : S (la saison), M (la météo), C (le climat).

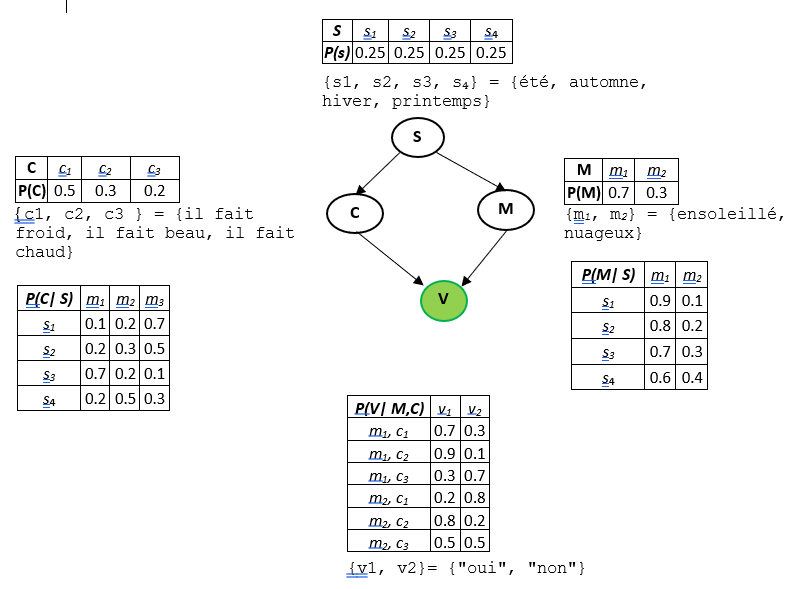


Figure . Exemple d’un RB à variables discrètes.

Les noms des variables sont comme suit S : Saison, C : Climat, M : Météo, V : Voyager.

### Condition de Markov

**Proposition1. (Pearl, 1988)** Soit un ensemble de variables, un graphe orienté acyclique sur et une distribution de probabilités sur. Le graphe représente si et seulement si chaque variable de est indépendante pour de tous ses non-descendants dans relativement à ses parents dans.

La structure d’un RB, définie par l’ensemble de ses arcs, exige que certaines contraintes d’indépendance doivent être vérifiées dans n’importe quelle distribution qui peut être représentée par. La propriété de Markov, énoncée dans la proposition, est une condition nécessaire et suffisante de représentation d'une distribution de probabilités par un graphe. En d’autres termes, la décomposition de la conjointe selon l’équation (II.9) n’est pas possible que lorsque chaque nœud est conditionnellement indépendant de ses non-descendants sachant ses parents directs dans le graphe. La condition de Markov suscite beaucoup de discussions contemporaines portant sur les RBs, elle a été utilisée pour la définition des RBs par plusieurs auteurs [6], [9].

Pour l’exemple de Figure II.1, parmi les indépendances conditionnelles qui découlent de la condition de Markov, on a les relations :,,,, ,, .

### Apprentissage des paramètres

Un RB est quantifié à l’aide d’un ensemble de probabilité conditionnelle*.* Dans un domaine discret, les tables de probabilité conditionnelle (TPC) sont souvent utilisées pour la modélisation de. La TPC correspondant à la variable est une table multidimensionnelle dont la taille est égale à (voir l’exemple de la Figure II .1). La case est la probabilité que la variable soit dans l’état étant donné que ses parents sont à l’état. L’apprentissage des valeurs de est une des tâches principales qui relèvent des RBs.

Dans cette section, nous allons décrire la méthode la plus utilisée dans le contexte des RBs.

#### Maximum de vraisemblance (MV)

Dans le cas où les données d’apprentissage sont abondantes et toutes les variables sont observées, la méthode MV est la plus utilisée pour l’estimation des tables de probabilités conditionnelles (TPCs). Les paramètres optimaux du RB sont ceux qui maximisent la vraisemblance des données d’apprentissage, , ici est une configuration pour l’ensemble des TPCs et est la configuration optimale exprimée comme suite :

*(‎II.10)*

Où est le nombre de classes (états), est la taille du vecteur d’attributs (features) etest le nombre d’événements dans la base de données pour lesquels la variable est dans l’état d’indice et ses parents sont dans la configuration d'indice. Vous trouvez la démonstration en [10].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | S=s1 | S=s2 | S= s3 | S= s₄ |
| C=c1 | 9 | 19 | 64 | 18 |
| C=c2 | 18 | 27 | 18 | 46 |
| C=c3 | 64 | 45 | 10 | 28 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | C=c1 | C=c2 | C= c3 |
| S=s1 | 9/91 | 18/91 | 64/91 |
| S=s2 | 19/91 | 27/91 | 45/91 |
| S= s3 | 63/91 | 18/91 | 10/91 |
| S= s₄ | 18/92 | 46/92 | 28/92 |

Figure . Exemple de MV.

Deux variables C =climat et S= saison. Le domaine de C est {c1=’il fait froid’, c2=’il fait beau’, c3=’li fait chaud’}. Le domaine de S est {s1=’été’, e2=’automne’, ’hiver’, ’printemps’}.

### Apprentissage de la structure

Un des problèmes complexes relatifs aux RBs est la construction du graphe d’un réseau à partir d’un échantillon de données d’apprentissage, par le mot "construction" nous voulons dire l’identification de l’ensemble d’arcs qui composent le graphe. L’apprentissage de la structure présente un compromis : comment peut-on construire un réseau dont la taille est minimale (un graphe moins dense) et qui représente la distribution conjointe?

Un réseau causal dont chaque arc *X→Y* signifie que *X* cause *Y* est un RB, l’inverse n’est pas toujours correct. Donc, il est préférable que la structure du graphe du RB soit donnée par l’expert en se basant sur les relations de causalité entre variables, malheureusement, cela n’est pas toujours possible et dans ce cas l’utilisation des algorithmes d’apprentissage automatique de la structure à partir de données devient une obligation.

La difficulté d’apprentissage de la structure des RBs peut être expliquée par les deux points suivants :

D’une part, l’indépendance (ou dépendance) statistique conditionnelle est un concept facile à définir en théorie, mais la confirmation d’une telle indépendance entre un ensemble de variables aléatoires sur la base d’un échantillon empirique d’observations est une tâche très délicate et sujette aux erreurs d’estimation.

D’autre part, le nombre de graphes possibles (espace de recherche) qui peuvent être formés à partir d’un ensemble de variables est super-exponentiel. Trouver la structure optimale c’est comme chercher une aiguille dans une botte de foin. Pour avoir une idée sur le gigantisme de l’espace de recherche, on peut former à partir de 4, 5, 10, 20, 30 et 40 nœuds, respectivement, 453, 29281, 4.7×1017, 2.34×1072, 2.71×10158 et 1.12×10276 graphes DAG différents [6].

### Algorithmes à base de tests d’indépendance conditionnelle

Les méthodes qui suivent cette approche commencent par la détermination à partir de données d’apprentissage les relations d’indépendance conditionnelle entre deux variables quelconques conditionnellement à un autre ensemble de variables. Ensuite, elles exploitent les résultats des tests statistiques d’indépendance pour en construire le graphe du RB d’une manière incrémentale. Les tests d’hypothèse d’indépendance de Pearson’s Chi-Square et de l’information mutuelle sont les plus utilisés. Parmi les méthodes populaires suivant cette approche, nous citons les algorithmes IC et IC\* de Pearl et Verma [8], les algorithmes SGS, PC, CI, FCI de Spirtes, Glymour et Scheines [5], et l’algorithme BN-PC de Cheng et al. [6],[7]. Souvent, ces algorithmes réalisent grossièrement les deux étapes suivantes :

1. Effectuer un ensemble suffisamment important de tests d’indépendance statistique conditionnelle.
2. Construire un graphe non orienté à partir des résultats des tests obtenus en étape.

### Inférence

#### Etat de l’art des algorithmes d’inférence

Le but de l’inférence dans les RBs est de calculer la marginale d’une variable d'intérêt sachant que nous disposons des valeurs réelles prises par d’autres variables, dites des évidences. Formellement, nous voulons évaluer et qui peut être réécrit en utilisant la règle de Bayes comme suit :

*(‎II.11)*

L’inférence sur les RBs est un problème pur du calcul. Le grand défi dans (II.11) est comment évaluer le numérateur d’une manière efficace. Ce dernier ne peut pas être évalué que par la marginalisation de la conjointe sur les autres variables. Pour ce faire, la solution triviale qui consiste à calculer toutes les valeurs de la conjointe (par force brute) n’est pas une solution souhaitable à cause de la complexité exponentielle en mémoire et en temps, mais plutôt il faut investir dans la structure du graphe et opter pour le calcul local (*Local computation)* [11],[12] en définissant des fonctions locales qui impliquent un nœud et ses voisins (conjointes locales) et une stratégie de propagation.

Les méthodes d’inférence exacte peuvent être scindées en deux classes :

1. Elimination de variables, lors de la marginalisation, certaines variables peuvent être éliminées en évaluant des conjointes locales de petites dimensions, l’ordre de l’élimination des variables dépend de la requête d’inférence et de la structure du graphe. A titre d’exemple, si nous voulons calculer la marginale de la variable sur le RB défini par le graphe, on a :

*(‎II.12)*

1. Algorithmes de propagation de messages sur le RB, qui utilisent un mécanisme d’échange de messages le long des arcs du réseau. Le premier algorithme, proposé par Pearl [16] en 1982 avant même le mûrement du concept des RBs, est restreint au cas des RBs dont le graphe est un arbre, une extension de cet algorithme aux graphes multi-arbres (Polytrees) est faite par Kim et Pearl [17]. Par la suite, des techniques expliquant comment cet algorithme peut être utilisé pour des graphes plus généraux qui peuvent avoir des boucles (cycles non orientés) ont été proposées par Pearl dans [4].

## Réseaux Bayésiens naïfs (RBNs)

### Définition du RBN

La classification Bayésienne naïve est un type de classification Bayésienne probabiliste simple basée sur le [théorème de Bayes](http://fr.wikipedia.org/wiki/Th%C3%A9or%C3%A8me_de_Bayes) dédiée au cas où il y a une forte indépendance (dite naïve) des hypothèses (des attributs). Elle met en œuvre un classificateur Bayésien naïf, appartenant à la famille des [classificateurs linéaires](http://fr.wikipedia.org/wiki/Classifieur_lin%C3%A9aire).

Un terme plus approprié pour le modèle probabiliste sous-jacent pourrait être « modèle à caractéristiques statistiquement indépendantes ».

En termes simples, un classificateur Bayésien naïf (CBN) suppose que l'existence d'une caractéristique (attribut) pour une classe est indépendante de l'existence d'autres caractéristiques. Par exemple, un fruit peut être considéré comme une pomme s'il est rouge, arrondi, et fait une dizaine de centimètres. Même si ces caractéristiques sont liées dans la réalité, un classificateur CBN déterminera que le fruit est une pomme en considérant indépendamment ces caractéristiques de couleurs, de forme et de taille.

Selon la nature de chaque modèle probabiliste, les CBNs peuvent être entraînés efficacement dans un contexte d'[apprentissage supervisé](http://fr.wikipedia.org/wiki/Apprentissage_supervis%C3%A9). Dans beaucoup d'applications pratiques, l'estimation des paramètres pour les modèles Bayésiens naïfs repose sur le [maximum de vraisemblance](http://fr.wikipedia.org/wiki/Maximum_de_vraisemblance) puisqu’ils comportent peu de paramètres à estimer ; des tables de probabilités conditionnelles en dimension 2.

Bien que les modèles CBNs ont une structure « naïve » et leurs hypothèses de base sont relativement simples, les CBNs ont fait preuve d'une efficacité plus que suffisante dans beaucoup de situations réelles complexes. En 2004, un article a montré qu'il existe des raisons théoriques derrière cette efficacité inattendue [19]. Toutefois, une autre étude en 2006 montre que des approches liées aux CBNs ([arbres renforcés](http://fr.wikipedia.org/w/index.php?title=Arbres_renforc%C3%A9s&action=edit&redlink=1), [forêts aléatoires](http://fr.wikipedia.org/wiki/For%C3%AAts_al%C3%A9atoires)) permettent d'obtenir de meilleurs résultats [20].

L'avantage du CBN est qu'il requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires à la classification, à savoir moyennes et variances des différentes variables. En effet, l'hypothèse d'indépendance des variables permet de se contenter de la variance de chacune d'entre elles pour chaque classe, sans avoir à calculer la [matrice de covariance](http://fr.wikipedia.org/wiki/Matrice_de_covariance).

### Structure (réseau Bayésien naïf)

Le modèle RBN est le modèle probabiliste conditionnel suivant :

*(‎II.13)*

Oùest une variable-classe dépendante dont les instances ou*les* classes possibles sont peu nombreuses, conditionnées par plusieurs variables caractéristiques*.*

Le RBN est un RB simple, sa structure comporte seulement des arcs qui partent de la variable classe vers les variables [18], voir l’exemple de la Figure II.3.

Bien que les RBNs soient assez simples, ils se sont montrés très efficaces dans de nombreuses applications. Ils sont recommandés lorsque les variables sont mutuellement indépendantes.

…..

…..

Figure II.3 Structure d’un réseau Bayesien naif.

### Classificateur Bayésien naïf

Un CBN () est un classifieur probabiliste basé sur un RBN dont la variable classe est le parent de toutes les variables} comme le montre la figure II.3, dans lequel la prédiction de la classe d’un vecteur de données est donnée par la fonction de décision suivante :

*(‎II.14)*

est la probabilité postérieure de la classe pour un objet étant donné le vecteur d’attributs .

### Hypothèses du modèle RBN

Les RBNs supposent une indépendance conditionnelle entre les variables ou les attributs ; c’est dire si . Dans le cas contraire, si cette hypothèse n’est pas vérifiée, alors théoriquement d’autres structures des RBs plus complexes que celle du CBN sont nécessaires.

### CBN avec données complètes

Soit un vecteur à classifier de dimension*,* et les classes possibles. Dans le cas où les valeurs des attributs sont toutes disponibles. La fonction (II.14) permet d’inférer la classe la plus probable pour le vecteur . Par l’exploitation de la propriété de l’indépendance conditionnelle des attributs, , la formule (II.14), peut être simplifiée en une forme décomposable pour donner la formule *(II.16).*

*(‎II.15)*

*(‎II.16)*

Dans (II.15) le dénominateur est constant et la factorisation en un produit dans (II.16) est une conséquence directe de l’indépendance conditionnelle des attributs.

### CBN avec données manquantes

Un grand avantage des RBNs est leur capacité immédiate à traiter le problème des données manquantes. Soit les attributs d’un vecteur à classifier. Lorsque les valeurs d'un sous-ensemble de, par exemple, sont inconnues ou manquantes, la fonction de classification peut faire la prédiction en se basant seulement sur les attributs disponibles, peut être obtenue immédiatement comme suit :

*(‎II.17)*

## Conclusion

Ce chapitre nous a permis d'introduire brièvement les RBs. Nous avons mis en évidence que ce type d'outils permet de modéliser les relations d’indépendance statistique conditionnelle entre les variables d’un problème d’application et qu’il permet également une manipulation formelle de l’incertitude de connaissances via le calcul probabiliste. Un des points forts des RBs est le fait qu’ils disposent d’une lisibilité certaine comparés à d'autres paradigmes tels que les réseaux de neurones.

On dénombre plusieurs applications des RBs dans des domaines variés où l’incertitude joue un rôle important. En effet, on trouve des applications en : bio-informatique et prise de décision médicale, diagnostique de défauts et analyse de fiabilité de systèmes, où la précision est un objective crucial.

Les RBNs représentent une variante des RBs disposant d’une structure simple. Ils sont recommandés pour la classification lorsque les attributs sont conditionnellement indépendants étant donné la variable de classe. Dans notre travail, nous allons utiliser les CBNs pour décider si une imagette (segment) correspond à un caractère, bruit ou un chevauchement qui nécessitera encore une autre opération de segmentation.

[1]. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition:  
a review. IEEE Trans Pattern Anal Mach Intell 22(1):4–37

[2]. [Nasser M. Nasrabadi](https://www.spiedigitallibrary.org/profile/Nasser.Nasrabadi-6820), "Pattern Recognition and Machine Learning," Journal of Electronic Imaging 16(4), 049901 (1 October 2007). <https://doi.org/10.1117/1.2819119>

[3]. Duda, R.O. and Hart, P.E. and Stork, D.G “Pattern Classification” Wiley 9781118586006 (2012). <https://books.google.dz/books?id=Br33IRC3PkQC>

[4]. J. Pearl, "Probabilistic reasoning in intelligent systems", Morgan Kaufman, San Mateo (Californie), 1988.

[5]. P. Spirtes, C. Glymour, et R. Scheines, "Causation, prediction, and search", Springer-Verlag, 1993.

[6]. J. Cheng, D. Bell, et W. Liu, "An algorithm for bayesian network construction from data", In Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics AISTAT’97, pages 83–90, 1997.

[8]. J. Pearl, "Causality: Models, Reasoning, and Inference", Cambridge University Press, England, 2000.

[7]. J. Cheng, D. Bell, et W. Liu, "Learning belief networks from data: An information theory based approach", In Proceedings of the sixth ACM International Conference on Information and Knowledge Management CIKM, pages 325–331, 1997.

[8]. J. Pearl, "Causality: Models, Reasoning, and Inference", Cambridge University Press, England, 2000.

[9]. J. Williamson, "Bayesian nets and causality", Oxford University Press, New York, 2005.

[10]. P. Naim, P.H. Wuillemin, P. Leray, O. Pourret et A. Becker, "Réseaux bayésiens", Eyrolles, Paris, 2004.

[11]. F. V. Jensen, S. L. Lauritzen et K.G. Olesen, "Bayesian, updating in causal probabilistic networks by local computations", Computational Statistics Quaterly, vol. 4, pp. 269–282, 1990.

[12]. S. Lauritzen et D. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems", Journal of the Royal Statistical Society, Series B, vol. 50, no. 2, pages 157–224, 1988.

[13] R. Stefan Niculescu, T. M. Mitchell et R. B. Rao, "Bayesian Network Learning with Parameter Constraints", Journal of Machine Learning Research, vol. 7, pages 1357–1383, 2006.

[14]. G. F. Cooper et E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data", Machine Learning, vol. 9, pages 309–348, 1992.

[15]. L. M. de Campos, J. M. Fernández-Luna et J. M. Puerta, "An iterated local search algorithm for learning Bayesian networks with restarts based on conditional independence tests", International Journal of Intelligent Systems, vol. 18, pages 221–235, 2003.

[16]. J. Pearl, ["Reverend Bayes on inference engines: A distributed hierarchical approach"](https://www.aaai.org/Papers/AAAI/1982/AAAI82-032.pdf), Proceedings of the Second National Conference on Artificial Intelligence. [AAAI-82: Pittsburgh, PA](http://www.aaai.org/Library/AAAI/aaai82contents.php). Menlo Park, California: AAAI Press, pages 133–136, 1982.

[17]. J.H Kim, J. Pearl, ["A computational model for combined causal and diagnostic reasoning in inference systems"](http://ijcai.org/Past%20Proceedings/IJCAI-83-VOL-1/PDF/041.pdf), Proceedings of the Eighth International Joint Conference on Artificial Intelligence, pages 190–193, 1983.

[18]. N. Friedman, D. Geiger, et M. Goldszmidt, "Bayesian network classifiers", Machine Learning, pages 131-161, 1997.

[19]. M. Mozina, J. Demsar, M. Kattan, et B. Zupan, "Nomograms for Visualization of Naive Bayesian Classifier", In Proc. of PKDD, 2004.

[20] R. Caruana, et A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms", Proceedings of the 23rd international conference on Machine learning, 2006.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **x1** | **x2** | **…** | **xd** | **etiq** |
| **X1** |  |  |  |  |  |
| **X2** |  |  |  |  |  |
| **…** |  |  |  |  |  |
| **Xn** |  |  |  |  |  |

**(a)**

**Fig. 2** la forme générale d’un dataset d’apprentissage

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** | **x4** | **…** | **xd** |
| **X** | ? |  |  | ? |  |  |

**(b)**

**Fig. 3** exemple d’un échantillon d’essai à prédire