

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOULOU MAMMERI, TIZI-OUZOU



FACULTE DE GENIE ELECTRIQUE ET D'INFORMATIQUE
DEPARTEMENT D'AUTOMATIQUE

MEMOIRE DE MAGISTER

en Automatique

Option : **Traitement d'Images et Reconnaissance de Formes**

Présenté par

MAZARI Chahrazad

Ingénieur U.M.M.T.O

Réseaux bayésiens appliqués à une fusion de données multi-capteurs en vue d'un contrôle intelligent d'un microclimat.

Mémoire soutenu le : 18/02/2015

devant le jury d'examen composé de :

LAGHROUCHE Mourad, Professeur à l'U.M.M.T.O **Président**
DIAF Moussa, Professeur à l'U.M.M.T.O **Rapporteur**
MELLAH Rabah, M.C.A. à l'U.M.M.T.O **Examineur**
BENSIDHOUM Mohand-Outahar, M.C.A. à l'U.M.M.T.O **Examineur**
ALKAMA Sadia, M.C.B. à l'U.M.M.T.O **Examineur**

Sommaire

Introduction générale

Chapitre 1 : Fusion de données multi-capteurs

1.1. Introduction	10
1.2. Fusion multi-capteurs	11
1.3. Méthodes de fusion	12
1.3.1 Méthodes basées sur des modèles physiques	13
1.3.2 Méthodes basées sur les techniques de l'IA	14
1.3.2.1 Théorie des probabilités	14
1.3.2.2 Théorie de l'évidence	16
1.3.2.3 Théorie des possibilités	19
1.4. Processus de fusion	20
1.5. Conclusion	24

Chapitre 2 : Les réseaux bayésiens

2.1. Introduction	25
2.2. Rappels sur les probabilités	25
2.3. Réseau bayésien	26
2.4. Apprentissage dans les réseaux bayésiens	28
2.4.1 Apprentissage des paramètres	28
2.4.2 Apprentissage de la structure	32
2.5. L'inférence	33
2.5.1 Les méthodes d'inférence exacte	34

2.5.2 Les méthodes d'inférence approchée	37
2.6. Réseaux bayésiens particuliers	38
2.6.1 Réseaux bayésiens dynamiques	39
2.6.2 Diagrammes d'influence	40
2.6.3 Classificateurs bayésiens	40
2.7. Conclusion	42

Chapitre 3 : Application des RB à une serre agricole

3.1. Introduction	43
3.2. Paramètres du microclimat dans une serre	44
3.2.1 Influence de la température	45
3.2.2 Influence de l'humidité	45
3.2.3 Influence de la lumière	46
3.3. Réseaux multi-capteurs dans une serre agricole	46
3.4. Méthodes de contrôle des serres agricoles	48
3.4.1 Méthode classique	48
3.4.2 Méthodes modernes	49
3.5. Application des réseaux bayésiens à la serre	53
3.5.1 Création de la base de données simulée	53
3.5.2 Outil de programmation	55
3.5.3 Création du réseau bayésien	59
3.5.3.1 Modélisation	59
3.5.3.2 Estimation des paramètres	60
3.5.3.3 Combinaison	61
3.5.3.4 Décision	63
3.6. Evaluation du modèle développé	64
3.6.1 Tests sur la base d'apprentissage	65

3.6.2 Tests sur la base de test	66
3.7. Conclusion	67
Conclusion générale	68
Références bibliographiques	70

Liste des tableaux

- Tableau 1 : Exemple de données numériques d'une variable continue.
- Tableau 2 : Les valeurs optimales des caractéristiques climatiques pour un olivier sous serre.
- Tableau 3 : Les différents états des paramètres climatiques.
- Tableau 4 : Les différentes classes du microclimat.
- Tableau 5 : Les valeurs de températures d'une région méditerranéenne de 24H.
- Tableau 6 : Exemple de deux classes du microclimat.
- Tableau 7 : Calcul des moyennes et des variances.
- Tableau 8 : Une nouvelle observation à classer.
- Tableau 9 : Résultats obtenus en appliquant l'algorithme de l'arbre de jonction sur la base d'apprentissage.
- Tableau 10 : Résultats obtenus sur la base de test de 54 observations.
- Tableau 11 : Résultats obtenus sur la base de test de 162 observations.

Liste des figures

Chapitre 1 Fusion de données multi-capteurs

1. Exemple d'un arbre de décision

Chapitre 2 Les réseaux bayésiens

1. Réseau bayésien simple
2. Exemple de réseau bayésien à 6 nœuds
3. Graphe de la variable continue X_1
4. Algorithme EM d'estimation des paramètres θ
5. Propagation par messages locaux dans un graphe
6. Réseau Bayésien à cinq nœuds
7. Graphe non dirigé dit moralisé
8. Graphe moral triangulé
9. Arbre de jonction
10. Réseau bayésien dynamique
11. Exemple d'un diagramme d'influence
12. Réseau bayésien naïf
13. Réseau bayésien augmenté par un arbre
14. Réseau bayésien multi-nets

Chapitre 3 Application des RB à une serre agricole

1. Multi-capteurs dans une serre agricole
2. Réseau Bayésien naïf des paramètres de la serre crée par Matlab
3. Représentation de la température climatique d'une région Méditerranéenne
4. Notations utilisées dans le réseau bayésien

Avant-propos

Je tiens en tout premier lieu à remercier mon directeur de mémoire Monsieur DIAF Moussa, professeur à l'université Mouloud MAMMERY de Tizi-Ouzou, pour la confiance qu'il m'a accordée en me proposant ce thème et pour m'avoir dirigée, conseillée et aidée tout le long de mon travail.

Nous ne manquerons pas de remercier Monsieur LAGHROUCHE Mourad, professeur à l'université Mouloud MAMMERY de Tizi-Ouzou pour m'avoir fait l'honneur d'accepter de présider le jury de ce mémoire.

Nous exprimons nos sincères remerciements à Monsieur MELLAH Rabah, Maître de conférences classe A à l'université Mouloud MAMMERY de Tizi-Ouzou pour avoir accepté de faire partie du jury d'examen de notre mémoire.

Nos sincères et vifs remerciements s'adressent aussi à Madame ALKAMA Sadia, Maître de conférences classe B à l'université Mouloud MAMMERY de Tizi-Ouzou qui a accepté d'examiner notre travail.

Nous exprimons également nos vifs remerciements à Monsieur BENSIDHOUM Mohand-Outahar, Maître de conférences classe A à l'université Mouloud MAMMERY de Tizi-Ouzou pour sa présence parmi les membres de jury.

Enfin, je remercie du fond du cœur et avec un grand amour mes parents qui n'ont jamais cessé de croire en moi pendant toutes mes années d'études, ma sœur, mon frère et mon cher époux qui m'ont toujours encouragée à y aller de l'avant. Un grand merci aussi à mes beaux-parents, à toute ma famille et mes amies.

Introduction générale

Dans différents environnements, souvent, la prise de décision n'était effectuée qu'après le recueil d'informations relevées par un agent humain à partir d'instruments, d'appareils ou de capteurs. Ces informations recueillies peuvent être incertaines, imprécises, incomplètes et entachées de subjectivité. De plus, cette tâche reste très contraignante et pénible et parfois à risques dans certains domaines. Ainsi, l'automatisation complète de la procédure s'avère nécessaire. Ceci conduit à une prise de décision objective en utilisant des moyens matériels permettant de vérifier et d'analyser rapidement et d'une manière autonome les informations. Ces systèmes s'appuient sur différentes théories basées sur des techniques de l'intelligence artificielle et de l'automatique, des probabilités, de décision, de la théorie des graphes, de la logique floue, etc. Les réseaux bayésiens auxquels nous nous intéressons dans ce présent travail constituent un modèle qui hybride la théorie des probabilités à celle des graphes et ce, pour prendre en charge la notion d'incertitude et d'incomplétude des données et qui trouvent une grande place dans plusieurs contextes applicatifs comme le diagnostic, la simulation du comportement des systèmes, l'analyse des données, la prise de décision, le contrôle des systèmes, etc. Dans notre cas, ces réseaux bayésiens sont appliqués dans le cadre d'une fusion de données multi-capteurs pour le contrôle automatique du microclimat dans une serre agricole. Initialement connues comme de simples abris contre les intempéries, les serres agricoles sont progressivement devenues de véritables systèmes industriels offrant une forte augmentation des rendements et une bonne qualité des produits

grâce à la prise en charge du contrôle et de la surveillance des paramètres de température, d'humidité, d'éclairement etc. En général, le processus de décision à partir d'une fusion multi-capteurs dans une application donnée tel que nous l'envisageons dans notre travail, peut être décomposé en deux étapes. Dans la première, les informations disponibles sont modélisées dans un cadre théorique permettant de prendre en compte les connaissances sur le phénomène étudié. Dans la seconde, une décision est prise en fonction de toutes les informations précédemment fusionnées.

Ainsi, le travail présenté dans ce mémoire est structuré en trois chapitres.

Dans le premier chapitre, nous présentons d'abord, la fusion multi-capteurs puis les grands principes théoriques de la fusion de données parmi lesquelles la théorie des probabilités, la théorie de l'évidence et celle des possibilités, ce qui nous amènera au processus de fusion.

Le second chapitre est consacré aux réseaux bayésiens. Nous présentons d'abord les différents algorithmes d'apprentissage et d'inférence les plus utilisés puis les réseaux bayésiens particuliers parmi lesquels les réseaux bayésiens naïfs qui font l'objet de notre application.

Dans le troisième chapitre, nous abordons quelques notions générales sur les paramètres climatiques, leurs influences sur les plantes, les capteurs utilisés pour les mesurer et leurs emplacements dans la serre. Nous passons, en revue, les différentes méthodes de contrôle de systèmes avec et sans modèles. Nous présentons, ensuite, les étapes suivies pour créer le système de contrôle considéré sous forme d'un problème de classification automatique tel que les classes correspondent aux différents états du climat et les formes représentent les paramètres du climat que sont la température ambiante, la température du sol, l'humidité ambiante et le rayonnement. Les deux étapes importantes de cette méthode sont la création de la base de données et l'application des algorithmes d'apprentissage et d'inférence pour classer une nouvelle mesure.

Ce mémoire est terminé par une conclusion générale et des perspectives ainsi que des références bibliographiques.

Chapitre 1

Fusion de données multi-capteurs

1.1. Introduction

La problématique de la fusion de données multi-capteurs peut prendre des chemins différents selon le contexte applicatif. Nous la retrouvons, en particulier, dans le domaine militaire, médical, la robotique etc. Quoique ces domaines sont différents, le principe de la fusion multi-capteurs reste approximativement le même pour la perception de l'environnement à partir de données issues de plusieurs sources en vue d'un éventuel contrôle, de l'établissement d'un diagnostic ou autre. L'utilisation de plusieurs sources simultanément est importante car une seule source permet, dans la majorité des cas, de n'avoir qu'une décision partielle sur l'environnement analysé. Ainsi, si plusieurs capteurs identiques sont utilisés, combiner les observations provoque des estimations améliorées de la mesure observée. Cependant, le même résultat peut être obtenu par la combinaison d'un certain nombre d'observations issues d'un même capteur à des instants différents [1]. On peut, également, combiner des données issues de capteurs mesurant des grandeurs différentes pour un instant précis dans le but d'évaluer l'état

du système observé. Dans ce cas, on peut intégrer la notion d'apprentissage où l'utilisation de la base de données est, justement, de faire intervenir des observations prises à des instants différents, sur une période de temps bien déterminée selon la nature des phénomènes étudiés et ce, dans le but, d'estimer les paramètres du modèle utilisé. Le but d'une telle approche, est la reconstitution du modèle de l'environnement observé à partir des signaux fournis par les capteurs et de la prise en compte des connaissances des experts du domaine. Toutefois, il est important de noter que fusionner des données n'est intéressant que si cela permet d'augmenter la qualité des prévisions et des décisions prises. Si les décisions prises sont de mauvaise qualité, elles peuvent avoir un impact négatif sur les résultats attendus d'une application. Cependant, l'amélioration de la qualité de la précision dans les mesures est liée à la variabilité des données obtenues en sortie par rapport aux données initiales du processus de fusion et dans le cas où les données initiales sont bruitées ou entachées d'erreurs, le processus de fusion de données tentera de réduire ce bruit et d'éliminer ces erreurs. Ainsi, dans ce qui suit, nous présentons la fusion multi-capteurs, les méthodes de fusion et le processus de fusion.

1.2. Fusion multi-capteurs

De nombreuses définitions ont été données à la fusion multi-capteurs selon les chercheurs. D'après D. Bellot dans [2], « *La fusion de données vise à l'association, la combinaison, l'intégration et le mélange de multiples sources de données représentant des connaissances et des informations diverses dans le but de fournir une meilleure décision par rapport à l'utilisation séparée des sources de données* ». Pour I. Blosh dans [3], « *La fusion d'informations consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décisions* ». Quant à Roger Reynaud [3], « *La fusion de données décrit les méthodes et les techniques numériques permettant de mélanger des informations provenant de sources différentes afin d'obtenir une décision* ». A partir de ces définitions, nous constatons que le sens de la fusion de données reste plus ou moins le

même dans le milieu des scientifiques et son cadre diffère selon les domaines d'application.

Le processus de fusion passe généralement par quatre principales étapes à savoir, la modélisation, l'estimation, la combinaison et la décision [4]. La modélisation est la première étape qui consiste à choisir le formalisme de la méthode à appliquer. Cette étape se base, en grande partie, sur les connaissances d'experts. Dans le contexte de la fusion de données multi-sources, chaque source fournit une information sur une décision candidate. Cette information peut être modélisée par une distribution de probabilités, une fonction, une formule ou tout autre selon le formalisme choisi. L'étape d'estimation suit, automatiquement, la modélisation et qui consiste à quantifier les informations des sources selon leurs natures, par exemple, sous forme de probabilités, si c'est le formalisme choisi. Pour la combinaison, il s'agit de regrouper les informations issues de différentes sources, en utilisant, un opérateur selon le formalisme choisi dans l'étape de modélisation. Le résultat, obtenu, sera de même nature et s'interprète de la même façon que les informations fusionnées. La décision constitue le cœur du processus de fusion, dont le but est la prise d'une décision plus élaborée que celle donnée par chacune des sources prise séparément. Le choix du critère se fait en fonction du formalisme. Généralement, il s'agit de la minimisation ou maximisation d'une fonction issue de la combinaison. Cette étape est sensée donner à l'utilisateur la meilleure décision possible.

1.3. Méthodes de fusion

Plusieurs méthodes de fusion ont été proposées dans la littérature. On rencontre, généralement, deux principales catégories de fusion qui sont celles basées sur des modèles physiques et celles basées sur l'utilisation des techniques de l'intelligence artificielle et la classification automatique qui peuvent se baser sur plusieurs théories comme la théorie des probabilités, la théorie de l'évidence, la théorie des possibilités et les sous-ensembles flous, on peut citer aussi les réseaux de neurones ainsi

que les Machines à Vecteurs Support (SVM) qui peuvent être utilisées comme des méthodes de fusion et de classification.

1.3.1. Méthodes basées sur des modèles physiques

Parmi ces modèles, le plus connu est le filtre de Kalman [5] qui est une méthode récursive de la fusion de données dont l'objectif est la prise en compte des erreurs de modélisation et celles des mesures en introduisant la notion de probabilités. Cette approche est largement utilisée pour le contrôle de systèmes dynamique dans un cadre de la fusion de données multi-capteurs. Le système peut être décrit par un modèle linéaire ou non linéaire. L'essentiel est qu'il soit connu à l'avance. L'erreur associée aux capteurs et au système est modélisée par un bruit gaussien. Dans le cas du filtre de Kalman linéaire, le problème d'estimation d'état du système se ramène à la résolution d'un système à deux équations: l'équation d'évolution du processus et l'équation de mesure. Ces équations sont la représentation d'état d'un système dynamique. Le modèle du processus qui décrit l'évolution de ce système, est défini par l'équation d'état linéaire. Quant au modèle de mesure, il décrit l'information fournie par le ou les capteurs est donné par une équation liant les paramètres de l'état, de la mesure et du bruit. Cette approche n'est applicable que lorsque les deux modèles d'évolution du système et celui de mesure sont linéaires. Dans le cas contraire, on fait appel à d'autres méthodes selon que le modèle d'évolution du système et/ou de mesure sont non linéaires. Pour le filtre de Kalman linéarisé, le modèle du système est non linéaire et le modèle de mesure est linéaire. On procède, alors, à une linéarisation autour de la trajectoire nominale pré-calculée pour pouvoir appliquer le filtre de Kalman linéaire. Dans le cas du filtre de Kalman étendu, la linéarisation, cette fois ci, est faite autour de la trajectoire estimée du filtre. La plus grande difficulté de cette méthode est l'établissement des modèles des systèmes dynamiques. Les méthodes de l'intelligence artificielle sont alors venues prendre place sans nécessiter un modèle formel du système étudié.

1.3.2. Méthodes basées sur les techniques de l'IA

En traitement de l'information, l'intérêt que représente une approche multi-sources dans de nombreux contextes d'applications est d'emblée évident. Comme exemple, prenons le cas de la fusion multi-capteurs dans le contexte de la classification. Le but est d'accéder à une information globale plus fiable et plus complète. La complémentarité et la redondance des informations sont deux facteurs très importants pour obtenir un tel effet. La complémentarité est beaucoup recherchée dans le domaine de la discrimination là où on a besoin de maximum de paramètres pertinents pour arriver à distinguer plusieurs situations avec plus de fiabilité et de sûreté. Cependant, la redondance permet d'avoir une meilleure précision en diminuant l'erreur sur les informations fournies par les sources. Cependant, et, contrairement, aux méthodes de la première catégorie, l'évolution de l'environnement observé n'a pas besoin d'être connu à l'avance et peut, souvent, être apprise. De nombreuses méthodes et techniques existent dans la littérature. Une bonne étude bibliographique nous a permis de regrouper les plus utilisées et qui présentent l'intersection entre l'intelligence artificielle et la fusion de données multi-capteurs utilisées dans le cadre de la classification. Parmi ces méthodes, celles issues de la théorie des probabilités, la théorie de l'évidence et la théorie des possibilités.

1.3.2.1. Théorie des probabilités

Dans le domaine de la fusion de données multi-capteurs, les méthodes probabilistes fusionnent les sorties des capteurs exprimées sous forme de probabilités, à la fin du processus de fusion, une décision est prise en fonction de toutes les informations fusionnées. Dans cette catégorie, on retrouve la régression logistique, l'arbre de décision, les k-plus proches voisins, les réseaux bayésiens, etc. Notons que la régression logistique est une technique de prédiction [6] dont le rôle est la construction d'un modèle permettant de prédire les valeurs prises par des variables cibles qualitatives Y (les classes), à partir d'un ensemble de variables caractéristiques X . C'est une technique qui produit un séparateur linéaire. La régression peut être binaire dans le cas où la

variable cible prend deux valeurs, ou polytomique dans le cas où elle possède plus de deux modalités. Si on se place dans le cas binaire $Y \in \{+, -\}$, la règle d'affectation qui est basée sur le théorème de Bayes nous permet de comparer les deux quantités respectives $P(Y = +/X)$ et $P(Y = -/X)$:

$$\frac{P(Y=+/X)}{P(Y=-/X)} = \left[\frac{P(Y=+)}{P(Y=-)} \times \frac{P(X/Y=+)}{P(X/Y=-)} \right] = R \quad (1.1)$$

Ainsi si le rapport $R > 1$ alors $Y = +$ sinon $Y = -$

Quant aux arbres de décision sont issues des travaux de Quinlan en 1986 puis en 1993 et qui représentent les techniques les plus connues et les plus utilisées en classification. C'est un processus récursif qui divise l'espace de données en sous régions de plus en plus proches des classes. Ils se composent de nœuds de décision contenant chacun une variable caractéristique, de branches correspondant chacune à l'une des valeurs possibles de la variable caractéristique sélectionnée et de feuilles qui sont les observations appartenant à la même classe (fig.1).

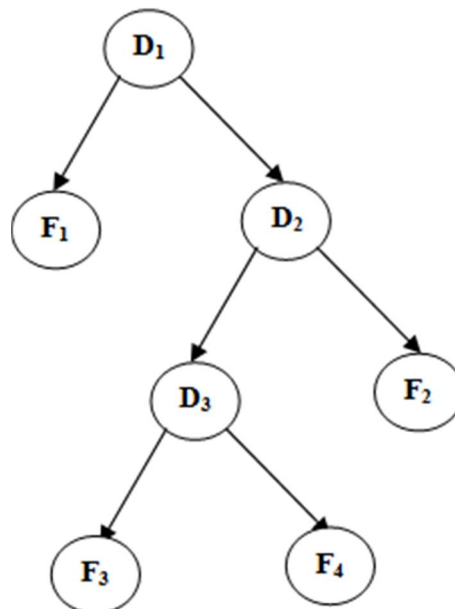


Fig.1 Exemple d'un arbre de décision

L'élaboration de l'arbre de décision passe par les mêmes étapes que les autres méthodes de classification, c'est-à-dire, la construction de la structure de l'arbre à partir d'un ensemble d'apprentissage puis

l'application de l'inférence, consistant à classer un nouvel exemple en suivant le chemin partant du nœud initial (D_1) à un nœud final (F_3), par exemple, en effectuant les différents tests pour chaque nœud de l'arbre. Parmi les algorithmes utilisés pour le faire, on retrouve : ID3 et C4.5 dans les travaux [7], [8] et [9] et CART dans [10]. Cependant, le principe de la méthode des k-plus proches voisins est la disposition d'une base d'apprentissage constituée d'un nombre suffisant de couples entrée-sortie (y, x_i) , ainsi pour estimer la sortie d'une nouvelle entrée, cette méthode prend en compte les k échantillons d'apprentissage dont l'entrée est la plus proche de la nouvelle entrée, selon une distance à définir comme la distance euclidienne :

$$d(y, x_i) = \sqrt{(y - x_i)^2} \quad (1.2)$$

Dans le cadre de la classification, nous retiendrons la classe la plus représentée parmi les k sorties associées aux k entrées les plus proches de la nouvelle entrée. Parmi les méthodes de la théorie de probabilités, on retrouve aussi les réseaux bayésiens qui font l'objet de notre application et qui seront présentés en détail dans le deuxième chapitre.

1.3.2.2. Théorie de l'évidence

De son côté, la théorie de l'évidence, nommée aussi le modèle de Dempster-Shafer ou la théorie des croyances, a été introduite par Shafer dans [11] en 1976 à la suite des travaux de Dempster sur les probabilités inférieure et supérieure [12] en 1967, en se basant sur des fonctions de croyance. C'est un formalisme qui permet de représenter, d'une part, l'imprécision et l'incertitude des informations issues de plusieurs capteurs par une approche non probabiliste à l'aide de fonctions de masse, de croyance et de plausibilité, et d'autre part, de combiner ces informations afin d'améliorer la prise de décisions.

Pour mettre en œuvre cette théorie, il est nécessaire de définir d'abord, l'espace de discernement et les différentes fonctions, pour ensuite arriver à l'estimation des fonctions de masse, la règle de combinaison et terminer par la règle de décision.

- *L'espace de discernement :*

Soit $D = \{C_1, C_2, C_3, C_4, \dots\}$ un univers, c'est-à-dire, un ensemble contenant toutes les décisions possibles pour un domaine donné. L'ensemble de ces parties $P(D)$ composé de 2^D éléments, est l'ensemble de tous les sous-ensembles de D , y compris l'ensemble vide. Par exemple si $D = \{a, b\}$, composé de deux éléments alors $P(D) = \{\emptyset, \{a\}, \{b\}, D\}$ est composé de $2^{|D|} = 2^2 = 4$ éléments. Les éléments de cet ensemble peuvent être interprétés comme des hypothèses. On peut, par exemple, interpréter l'élément $\{a\}$ comme « l'hypothèse a est vérifiée » ou « on est dans l'état a », ou encore $\{a, b\}$ comme « on est, soit dans l'état a ou dans l'état b ».

- *Fonctions de masse:*

La fonction de masse m représente la croyance accordée à un élément de l'ensemble de discernement, elle est définie comme étant une application de $2^{|D|}$ dans $[0,1]$.

$$m: 2^{|D|} \rightarrow [0,1] \quad (1.3)$$

La fonction de masse de l'ensemble vide est nulle et la somme des fonctions de masse des autres sous-ensembles vaut 1 :

$$\sum_{A \in P(D)} m(A) = 1 \quad (1.4)$$

La fonction de masse se différencie de la probabilité, par le fait, qu'elle attribue une croyance, pas seulement, aux singletons, comme a ou b , mais aussi aux composites de D comme D l'ensemble tout entier.

- *Fonctions de croyance et de plausibilité :*

A partir de la valeur de la fonction de masse d'un état, on peut définir un intervalle de confiance. Cet intervalle contient la valeur précise de la fonction de masse de l'état, bornée par deux mesures appelées croyance et plausibilité:

$$Bel(A) \leq P(A) \leq Pl(A) \quad (1.5)$$

La croyance $bel(A)$ d'un ensemble A représente la part totale de croyance soutenant A , elle est définie comme la somme des fonctions de masse de tous ses sous-ensembles :

$$Bel(A) = \sum_{B/B \subseteq A} m(B) \quad (1.6)$$

La plausibilité $pl(A)$ représente la part maximale de croyance qui pourrait soutenir A , elle est définie comme la somme des fonctions de masse de tous les ensembles B qui intersectent A :

$$Pl(A) = \sum_{B/B \cap A \neq \emptyset} m(B) \quad (1.7)$$

Ces deux mesures sont liées par la relation suivante : $Pl(A) = 1 - Bel(\bar{A})$

De ce fait, la connaissance d'une seule de ces valeurs, croyance ou plausibilité suffit pour déduire l'autre.

- *Estimation des fonctions de masse*

Cette étape peut faire appel à plusieurs règles [13]. Dans le modèle probabiliste, $m(\{C_i\})(x)$ est estimée comme une probabilité. Ce modèle peut être modifié soit par affaiblissement pour modéliser la fiabilité des sources : $m'(C_i) = \alpha m(C_i)$ et $m'(D) = 1 - \alpha + \alpha m(D)$ où α : est la fiabilité de la source par exemple soit par affaiblissement et pondération : $m^i(\{C_i\})(x) = \frac{\alpha_i R p(x/C_i)}{1 + \alpha_i R p(x/C_i)}$. Il existe bien d'autres règles d'estimations de ces fonctions, où R est une fonction de raffinement.

- *Règle de combinaison*

La règle de combinaison originale, connue en tant que règle de combinaison de Dempster, met clairement en valeur l'accord entre les différentes sources et masque les conflits grâce à un facteur de normalisation. La combinaison est calculée à partir des deux fonctions de masse m_1 et m_2 pour deux sources de la manière suivante :

$$m_{1,2}(\emptyset) = 0$$

$$m_{1,2}(A) = \frac{1}{1-k} \sum_{B \cap C = A \neq \emptyset} m_1(B) m_2(C) \quad (1.8)$$

où : $k = \sum_{B \cap C \neq \emptyset} m_1(B) m_2(C)$

k est une mesure du niveau de conflit entre les deux masses. Le facteur de normalisation $1 - k$ permet de masquer ces conflits et d'attribuer toute masse impliquée dans le conflit à l'ensemble nul.

- *Règle de décision*

Dans l'étape de décision, plusieurs règles sont possibles comme:

- le maximum de plausibilité : $x \in C_i$ si $Pl(C_i)(x) = \max\{Pl(C_k)(x), 1 \leq k \leq n\}$;
- le maximum de croyance : $x \in C_i$ si $Bel(C_i)(x) = \max\{Bel(C_k)(x), 1 \leq k \leq n\}$;

- le maximum de croyance sans recouvrement des intervalles de confiance : $x \in C_i$ si $Bel(C_i)(x) \geq \max\{Pl(C_k)(x), 1 \leq k \leq n, k \neq i\}$,
- les règles mixtes (plausibilité pour certaines classes et croyance pour d'autres).

1.3.2.3. Théorie des possibilités

La théorie des possibilités est une théorie qui a existé après la théorie des probabilités. Elle est issue de la théorie des sous-ensembles flous développée par Zadeh en 1965[14]. En 1978, la théorie des possibilités a été introduite par Dubois et Parade [15]. Cet outil mathématique permet de modéliser l'imprécision et l'incertitude dues à la variabilité des phénomènes qui nous entourent d'une part et l'incomplétude de l'information disponible d'autre part. Il s'agit ensuite de combiner des informations très variées grâce à la richesse et la souplesse des opérateurs utilisées. Cette méthode permet de décrire un événement par le degré de possibilité et le degré de nécessité. La valeur de ces deux grandeurs est comprise entre 0 et 1. Plus la mesure de possibilité est proche de 1, plus la réalisation de cet événement est possible. Si on prend deux événements contraires, dans la théorie des possibilités, au moins un événement est tout à fait possible telle que la possibilité de chacun des événements vaut 1. Par contre, dans la théorie des probabilités, la somme des probabilités des deux événements est équivalent à 1. Cette contrainte sur le fait que deux événements contraires ne peuvent pas se réaliser au même temps est levée par la mesure de nécessité de l'évènement, définie comme l'impossibilité de réalisation de l'évènement contraire pour un ensemble d'éléments S vérifiant un certain nombre de propriétés bien sélectionnées (caractéristiques pertinentes). La fonction caractéristique d'un sous-ensemble A de S est définie comme étant une fonction binaire qui vérifie l'appartenance d'un élément x de S à A tel que :

$$f_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} \quad (1.9)$$

Si $S = \{a, b\}$ et $A = \{a\}$ alors $f_A(a) = 1$ et $f_A(b) = 0$

Cependant, la théorie des sous-ensembles flous propose une fonction d'appartenance graduelle au lieu d'une fonction caractéristique binaire qui peut être considérée comme une fonction d'appartenance graduelle particulière. En termes de classification, cette fonction représente le degré d'appartenance d'une observation à une classe donnée ou encore le degré de possibilité d'une décision pour une observation donnée. Après cette étape, vient l'étape d'estimation de ces distributions de possibilités en utilisant plusieurs méthodes parmi celles utilisées en classification automatique et qui consiste à les déduire à partir de probabilités par plusieurs critères proposés dans [4]. Comme exemple, celui de la relation sur tous les sous-ensembles de A [16] telle que la probabilité de l'évènement A sera comprise entre la nécessité de A et sa possibilité. L'un des intérêts de la théorie des possibilités en fusion de données est donc la disponibilité d'un grand choix d'opérateurs de combinaison et leurs souplesses. Cependant, un classement doit être fait afin de choisir le meilleur opérateur à utiliser. Parmi les plus considérés, on retrouve les opérateurs conjonctifs, disjonctifs ou de compromis [17]. Une fois toutes les informations issues de différents capteurs sont combinées, la décision sera prise directement en prenant le maximum des degrés d'appartenance ou des distributions de possibilités.

1.4. Processus de fusion

La fusion de données a été identifiée comme étant un processus à trois niveaux fonctionnels d'après [18] :

- la fusion de données qui correspond à une acquisition d'informations directement depuis les capteurs,
- la fusion de caractéristiques qui correspond à la fusion de caractéristiques extraites, directement, des données issues des capteurs,
- la fusion de décisions liée à la fusion d'informations correspondant à la formulation d'hypothèses issues d'un expert ou d'un système comme pour le cas d'un classificateur.

La nature hypothétique de la fusion de données, est due, d'une part, à l'imprécision associée aux données initiales issues des différents capteurs et, d'autre part, à la capacité limitée de ces derniers de n'observer qu'une partie de l'environnement global considéré, ce qui rend nécessaire l'utilisation de mécanismes de contrôle sophistiqués basés sur l'intégration d'hypothèses multiples.

Ainsi, la fusion de données peut être vue comme un processus dont le but est de combiner l'information afin d'améliorer le processus de prise de décision comme dans le cas du contrôle d'un microclimat. Provenant d'un certain nombre de capteurs, la fusion de données forme un processus qui a pour but l'amélioration du processus de régulation des conditions climatiques désirées.

Les trois éléments fondamentaux de la fusion sont donc :

- les sources de données qui fournissent les ensembles initiaux de données qui seront fusionnés,
- le processus de fusion de données,
- les algorithmes qui lui sont associés,
- le résultat, qui est un ensemble de données lui-même.

Dans certains cas, les sources contiennent des données incertaines ou bruitées. Ce type de sources de données est particulièrement fréquent dans les problèmes de gestion de l'environnement pour lequel les capteurs électroniques sont, en effet, sujets au bruit et ne fournissent jamais une information sûre. Par exemple, pour augmenter les capacités de gérer une serre agricole, plusieurs types de capteurs sont spatialement distribués tels que des capteurs de température ambiante ou du sol, capteurs d'humidité, du degré de rayonnement, ou encore de taux de CO₂.

Les données fournies par une source peuvent être numériques ou symboliques, c'est à dire uniquement qualitatives. Objectives, il n'y a aucune incertitude associée à cette information. Il peut s'agir, par exemple, de faits exacts dans une base de connaissances, subjectives si les informations sont incertaines et bruitées. Ce dernier cas est le plus courant. En effet, il y a une différence notable entre les données réelles et les données observées fournies par la source de données, temporelles ou

causales lorsqu'il s'agit plus, ici, d'une caractéristique sur les relations entre les données que sur les données elles-mêmes. En effet, une source peut fournir des données qui sont toutes corrélées de façon temporelle, causale. Il existe, donc, un ordre temporel entre les données, ou encore une relation de cause à effet entre les données, les données peuvent aussi être spatiales quand dans l'espace de représentation des données, il existe une relation entre les données, mesurable par une métrique. Cependant, connaître les relations entre les sources est important, car, ceci permettra de caractériser, d'ordonner ou encore de corrélérer les différentes sources de données disponibles. On pourra, alors, décider quelle est la meilleure stratégie ou le meilleur algorithme pour fusionner les données.

Parmi les relations qui ont été identifiées dans [19], nous pouvons citer la distribution, la complémentarité, l'hétérogénéité, la redondance, la contradiction, la concordance, la discordance et la synchronisation.

Dans le cas de la distribution, les capteurs sont placés dans le même environnement mais chacun ayant soit un point de vue différent, soit la capacité de n'observer qu'une partie de l'environnement. Par exemple, dans une serre agricole, le nombre de capteurs utilisés pour le même paramètre est en relation avec la dimension de cette dernière, ainsi, chaque capteur nous renseigne, seulement, sur la région où il est placé.

Dans le cas de la complémentarité, chaque capteur perçoit uniquement un sous-ensemble de l'environnement global. La fusion de l'ensemble de ces capteurs donnera une vue élargie de l'environnement. Les sources, dans ce cas, ne sont pas redondantes mais complémentaires. C'est le cas d'une serre agricole, où la perception du climat dépend de l'utilisation de plusieurs capteurs, chacun d'entre eux nous renseigne sur un paramètre précis, température, humidité, etc.

Dans le cas de l'hétérogénéité, chaque source fournit des informations dont les caractéristiques sont complètement différentes des autres sources. Il s'agit là d'un cas très classique. Par exemple, toujours dans une serre, chacun des capteurs fournit des données météorologiques s'intéressant à différentes caractéristiques du climat voir température, humidité de l'air, rayonnement ou encore taux de CO₂ ou autre.

Dans le cas de la redondance, les sources de données décrivent le même environnement avec des informations de même caractéristique. C'est le cas lorsqu'un capteur est doublé afin de prévenir une panne ou simplement pour augmenter la précision ou la confiance accordée aux mesures. Si des différences apparaissent entre les mesures, alors un processus de fusion de données aura intérêt à utiliser ces différences pour améliorer la qualité du traitement de données et extraire de nouvelles informations. La redondance est nécessaire pour sécuriser ou rendre plus robuste un système percevant un environnement.

Dans le cas de la contradiction, l'espace de définition de deux sources est le même mais elles fournissent des informations totalement différentes l'une de l'autre, tout en observant le même environnement. Par exemple, un capteur de température pourra fournir comme mesure 10°C alors qu'un autre fournira 30°C, nous signalons que les deux capteurs sont placés dans le même environnement et dans les mêmes conditions.

Dans le cas de la concordance, les données fournies par deux sources sont compatibles entre elles et se confirment l'une l'autre. Par exemple, une première mesure issue d'un capteur de température fournira 25°C et une seconde mesure issue d'un deuxième capteur fournira la même mesure 25°C, cette mesure confirmera la première.

Dans le cas de la discordance, les informations fournies par deux sources sont incompatibles entre elles. La différence entre la contradiction et la discordance est qu'ici une source ne va pas invalider l'information donnée par l'autre source, mais simplement donnera un résultat qui est aussi acceptable que celui de la première source. Par exemple, un capteur de température peut fournir 25°C et un autre capteur 26°C. Si le domaine des capteurs se situe au moins dans l'intervalle alors les deux réponses, bien que discordantes, sont valides.

Dans le cas de la synchronisation, les données fournies par les sources sont concordantes au niveau du temps. La notion de temps et de date de délivrance d'une donnée est nécessaire. Par exemple, un capteur de température peut fournir des mesures chaque heure, alors qu'un autre fournira simplement des informations à la demande, la température la

matinée, l'après-midi et le soir n'est pas la même, c'est pour cela que la notion du temps est très importante.

Cependant, deux sources de données peuvent être concernées par plusieurs de ces caractéristiques à la fois. Par exemple, dans le cas de deux capteurs de température placés dans deux endroits différents de la serre, les deux capteurs observent le même environnement selon un point de perception différent, ces deux capteurs seront donc distribués et complémentaires. Ils contrôlent le même environnement (complémentaires) mais à partir de deux endroits différents (distribués). Ils ne sont pas redondants, car l'information fournie par chacun d'eux est différente puisque, ici encore, leurs points de vue sont différents. Ils ne sont pas discordants, sauf dans le cas d'un dysfonctionnement de l'un d'eux. Ils sont synchrones et sont normalement concordants.

1.5. Conclusion

Dans ce chapitre, nous avons pu voir l'intérêt, les fondements et les différents formalismes sollicités pour réaliser un système de décision par fusion de données issues de plusieurs capteurs. Nous avons introduit quelques méthodes utilisées dans le cadre de classification telles que les réseaux bayésiens, comme des fonctions de décision dans un système de contrôle à fusion multi-capteurs. Une étude détaillée des mécanismes de cette méthode fera l'objet du chapitre suivant. Notons, toutefois, qu'il existe d'autres méthodes qui peuvent être considérées comme des méthodes de fusion de données comme les chaînes de Markov cachées, les réseaux de neurones, les machines à vecteurs support, etc.

Chapitre 2

Les réseaux bayésiens

2.1. Introduction

Apparus pour la première fois en 1988 suite aux travaux de l'informaticien américain Judea Pearl (1936) et une équipe de chercheurs danoise dont l'objectif était d'introduire la notion d'incertitude dans les systèmes experts en utilisant le théorème de Bayes, les réseaux bayésiens sont des modèles de connaissances probabilistes basés sur l'apprentissage pour l'aide à la décision, le diagnostic ou le contrôle des systèmes complexes. Aujourd'hui, ils sont utilisés dans plusieurs domaines parmi lesquels le domaine militaire, médical et industriel. C'est une méthode qui marie la théorie des graphes à celle des probabilités. Leur principal avantage est la facilité de représentation des connaissances sous forme d'un graphe causale. Ainsi, dans ce chapitre, il s'agit, dans un premier temps, de définir cette méthode et passer à la description des différents algorithmes d'inférence et d'apprentissage ainsi que des classificateurs bayésiens.

2.2. Rappels sur les probabilités

Rappelons que la théorie des probabilités a pour objectif de modéliser des expériences pour lesquelles plusieurs résultats peuvent se réaliser mais leur réalisation ne peut pas être déterminée à l'avance.

Aussi, si un ensemble Ω a un nombre d'éléments fini N , alors $P(\Omega)$ l'ensemble des parties de Ω aura 2^N éléments. Si, par exemple, $\Omega = \{1,2\}$, $N = 2$ éléments, alors $P(\Omega)$ aura $2^N = 2^2 = 4$ éléments tel que $P(\Omega) = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$. Les événements considérés A sont des sous-ensembles de Ω : $A \subset \Omega$. Rappelons aussi que par définition, une probabilité P est une fonction de $P(\Omega)$ dans $[0,1]$:

$$\begin{aligned} P(\Omega) &\rightarrow [0,1] \\ A &\rightarrow P(A) \end{aligned} \quad (2.1)$$

tel que :

$$\begin{aligned} 1) &\forall A \subset P(\Omega), 0 \leq P(A) \leq 1 \\ 2) &P(\Omega) = 1 \\ 3) &P(A \cup B) = P(A) + P(B) \text{ si } A \text{ et } B \text{ sont indépendants.} \end{aligned} \quad (2.2)$$

Par ailleurs, on parle d'indépendance de deux événements A et B , si et seulement si :

$$P(A \cap B) = P(A) \times P(B) \quad (2.3)$$

$$P(A/B) = P(A) \quad (2.4)$$

$$P(B/A) = P(B) \quad (2.5)$$

De plus, A et B sont indépendants conditionnellement à un événement C si et seulement si :

$$P(A/B, C) = P(A/C) \quad (2.6)$$

On notera que $P(A, B) = P(A \text{ et } B)$ est la probabilité jointe alors que la probabilité conditionnelle est $P(A/B) = \frac{P(A, B)}{P(B)}$, définie seulement lorsque $P(B) \neq 0$.

De cette probabilité conditionnelle, le théorème de Bayes en découle tel que :

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} \quad (2.7)$$

La généralisation de ce théorème donne :

$$P(A/B) = \frac{P(B/A) \times P(A)}{\sum_A P(B/A) \times P(A)} \quad (2.8)$$

2.3. Réseau bayésien

Un réseau bayésien est représenté par un graphe dirigé sans cycles, composé de nœuds reliés par des arcs représentant une relation de cause (parents) à effet (enfants) (Fig.2.1).

Il est, par définition, un graphe orienté acyclique $G = (V, E, \Omega)$ où V est l'ensemble des nœuds représentant les variables du domaine considéré, E , l'ensemble des arcs représentant les relations qui existent entre les

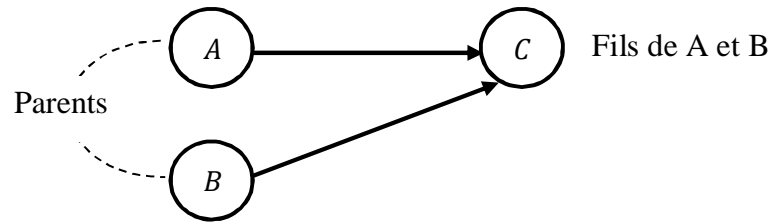


Fig.2.1 Réseau bayésien simple

différentes variables et $\Omega = P(x/p(x))$, l'ensemble des distributions de probabilités de chaque nœud x conditionnellement à ses parents immédiats $p(x)$ dans le réseau G .

Notons que si chaque nœud est conditionnellement indépendant de ses parents, il serait possible, alors, de factoriser la distribution conditionnelle par une probabilité jointe sur l'ensemble des nœuds V en faisant un produit comme suit :

$$P(V) = \prod_{x \in V} P(x/p(x)) \quad (2.9)$$

où $P(V)$ est la probabilité jointe globale et $P(x/p(x))$ la probabilité conditionnelle locale.

Comme exemple, prenons l'ensemble de nœuds $V = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ reliés par la structure de la figure 2.2. La probabilité jointe de l'ensemble des nœuds en fonction des probabilités conditionnelles est :

$$P(V) = P(x_6/x_5) P(x_5/x_2, x_3) P(x_4/x_1, x_2) P(x_3/x_1) P(x_2/x_1) P(x_1) \quad (2.10)$$

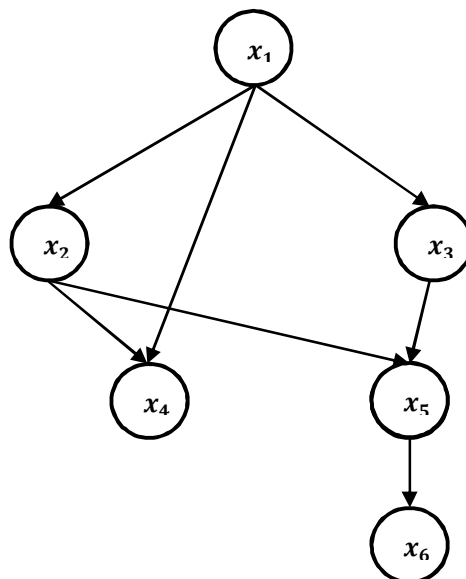


Fig.2.2 Exemple de réseau bayésien à 6 nœuds

Construire un réseau bayésien revient à définir le graphe ou structure du modèle et les probabilités ou paramètres de chaque variable conditionnellement à ses causes. En général, la structure est donnée par les experts du domaine considéré et les paramètres sont calculés à partir des données expérimentales. Le graphe du réseau peut être statique ou dynamique. L'intérêt particulier des réseaux bayésiens est de tenir compte, simultanément, des connaissances *a priori* d'experts et de l'information contenue dans les données.

2.4. Apprentissage dans les réseaux bayésiens

Soit G le graphe d'un réseau bayésien, D l'ensemble des données et θ l'ensemble de ses paramètres. $\theta = P(x_i/pa(x_i))$ est la probabilité de chaque nœud x_i conditionnellement à ses parents $pa(x_i)$. Le problème d'apprentissage consiste à faire l'apprentissage des paramètres et celui de la structure. Si la structure est connue, il faudra juste estimer les paramètres du réseau correspondant aux distributions de probabilités conditionnelles des variables. Dans le cas de l'apprentissage de la structure, les probabilités sont supposées connues et le but sera de trouver la meilleure connexion entre les variables d'intérêt du domaine considéré pour obtenir le meilleur graphe.

2.4.1. Apprentissage des paramètres

Les paramètres du réseau peuvent être donnés par un expert ou appris automatiquement à partir d'une base d'exemples. Plusieurs techniques sont envisageables selon la disponibilité des données d'après [20] et [21] et la nature de la variable qui peut être discrète ou continue. Dans le cas où toutes les variables sont observées, plusieurs approches existent. Nous présentons ici les plus utilisées dans le cadre des réseaux bayésiens. On parle d'apprentissage statistique dans le cas où les paramètres θ sont calculés de telle façon à maximiser $P(D/\theta)$. Le critère utilisé est le maximum de vraisemblance énoncé comme suit :

$$\theta^{MV} = \operatorname{argmax} P(D/\theta) \quad (2.11)$$

Cette approche consiste à estimer la probabilité d'un événement par la fréquence d'apparition de ce dernier dans la base d'exemples. Dans le cas où la variable concernée est discrète :

$$P(X_i = x_k / pa(X_i) = x_j) = \theta_{i,j,k}^{MV} = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}} \quad (2.12)$$

où $N_{i,j,k}$ est le nombre d'événements dans la base d'exemples pour lesquels la variable X_i est dans l'état x_k et ses parents sont dans la configuration x_j . L'aspect des variables continues dans l'apprentissage peut être pris en considération de deux façons différentes soit, en discrétisant la variable continue et appliquer les méthodes d'apprentissage discrètes, soit, en faisant une hypothèse sur la nature de la distribution. Ainsi, les paramètres du modèle ne seront plus les fréquences d'apparition comme nous l'avons vu, précédemment, mais, les paramètres de la loi de probabilité considérée. Dans ce cas, pour estimer ces paramètres, on fait appel aux distributions des lois continues de probabilités dont la plus adaptée pour modéliser des phénomènes naturels à aspect aléatoire, continu et dynamique est la loi gaussienne ou normale. Cette loi donnée par l'expression 2.13 a aussi l'avantage de modéliser l'imprécision qui, souvent, entache nos mesures.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.13)$$

Ainsi, lorsqu'une variable aléatoire X suit une loi normale, on a :

$$X \sim N(\mu, \sigma^2) \quad (2.14)$$

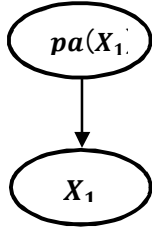
Ainsi, pour définir une loi normale à partir d'une base d'exemples, il suffit de définir la moyenne μ et l'écart-type σ . La probabilité conditionnelle d'un exemple donné sera donc estimée par :

$$P(X_i = x_k / pa(X_i) = x_j) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_k - \mu}{\sigma}\right)^2} \quad (2.15)$$

Comme exemple, soit le tableau 1.

Tableau 1 Exemple de données numériques d'une variable continue

La variable $pa(X_1)$	La variable X_1	μ	σ
1	21.6118	21.4947	2.5428
	17.7233		
	20.3208		
	24.7092		
	19.8557		
	23.9219		
	22.2481		
	24.8734		
	18.0782		
	21.6046		



Puisque X_1 suit une loi normale, alors $X_1 \sim N(21.4947, 2.5428^2)$ et le graphe est donné en figure 2.3.

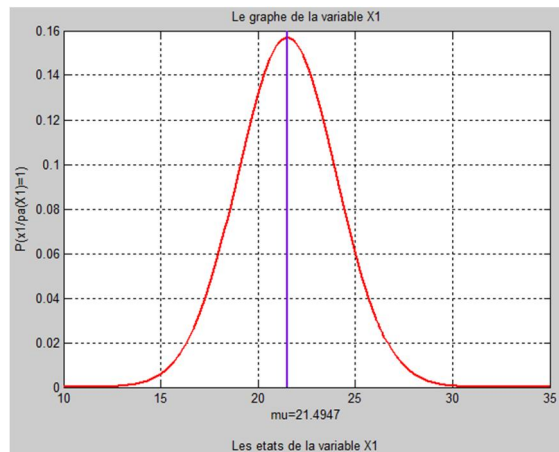


Fig.2.3 Graphe de la variable continue X_1

Ainsi, pour un exemple donné $X_1 = 19$, $P(X_1 = 19/pa(X_1) = 1) = 0.097$.

Le problème qui se pose en utilisant le maximum de vraisemblance est dans le cas où un exemple n'apparaît pas dans la base d'apprentissage. Sa fréquence est nulle. Ainsi, pour y remédier, on utilise l'apprentissage bayésien qui consiste à trouver les paramètres θ les plus probables sachant que les données ont été observées $P(\theta/D)$. Il s'agit du *maximum a posteriori* énoncé comme suit :

$$\theta^{MAP} = \operatorname{argmax} P(\theta/D) = \operatorname{argmax} P(D/\theta)P(\theta) \tag{2.16}$$

Ainsi, les paramètres satisfaisant cette règle sont donnés par la relation suivante :

$$P(X_i = x_k/pa(X_i) = x_j) = \theta_{i,j,k}^{MAP} = \frac{N_{i,j,k} + \alpha_{i,j,k} - 1}{\sum_k (N_{i,j,k} + \alpha_{i,j,k} - 1)} \tag{2.17}$$

Où $\alpha_{i,j,k}$ sont les paramètres de Dirichlet qui interviennent dans le calcul de la loi *a priori* $P(\theta)$.

Il existe, aussi, l'approche bayésienne, qui consiste à considérer l'espérance à la place du maximum. Il s'agit, alors, de l'espérance *a posteriori* qui nous permet de calculer ces paramètres par :

$$P(X_i = x_k / pa(X_i) = x_j) = \theta_{i,j,k}^{EAP} = \frac{N_{i,j,k} + \alpha_{i,j,k}}{\sum_k (N_{i,j,k} + \alpha_{i,j,k})} \quad (2.18)$$

où $\alpha_{i,j,k}$ sont les paramètres de Dirichlet.

Les données incomplètes peuvent parfois apparaître dans un réseau bayésien. Dans [22], elles sont classées suivant trois natures MCAR (*Missing Completely At Random*), MAR (*Missing At Random*) et NMAR (*Not Missing At Random*). Parmi les algorithmes les plus utilisés qui tentent d'estimer ces paramètres, l'algorithme EM (*Expectation-Maximisation*) apparu dans [23] et utilisé dans le cadre des réseaux bayésiens dans [24][25]. Cet algorithme composé de quatre étapes est itératif général d'estimation des paramètres θ qui peut être utilisé dans les deux cas d'apprentissage statistique et bayésien (Fig.2.4).

- 1- Initialiser les paramètres $\theta^{(0)}$ manquants.
- 2- **L'étape E** : Faire la somme de tous les paramètres $\theta_{i,j,k}$ dont l'état des variables et leurs parents est le même tel que :
Si l variant de 1 a N est le nombre d'exemples dans la base de données pour lesquels la variable X_i est dans l'état x_k et ses parents sont dans la configuration x_j , alors :

$$N_{i,j,k}^* = \sum_{l=1}^N P_l (X_i = x_k / pa(X_i) = x_j)$$
- 3- **L'étape M** : Consiste à mettre à jour les paramètres $\theta^{(t)}$ à partir des données complétées dans l'étape précédente, en utilisant les algorithmes d'apprentissage à partir de données complètes MV, MAP ou EAP.
 - Apprentissage statistique par maximum de vraisemblance

$$\theta_{i,j,k}^{(t+1)} = \frac{N_{i,j,k}^*}{\sum_k N_{i,j,k}^*}$$
 - Apprentissage Bayésien par maximum a posteriori

$$\theta_{i,j,k}^{(t+1)} = \frac{N_{i,j,k}^* + \alpha_{i,j,k} - 1}{\sum_k (N_{i,j,k}^* + \alpha_{i,j,k} - 1)}$$
 - Apprentissage Bayésien par espérance a posteriori

$$\theta_{i,j,k}^{(t+1)} = \frac{N_{i,j,k}^* + \alpha_{i,j,k}}{\sum_k (N_{i,j,k}^* + \alpha_{i,j,k})}$$
- 4- Application du critère d'arrêt : $|\theta^{(t+1)} - \theta^{(t)}| \leq \varepsilon$
Cela veut dire que tant que $\theta^{(t+1)} \neq \theta^{(t)}$, les itérations continues.

Fig.2.4 Algorithme EM d'estimation des paramètres θ

2.4.2. Apprentissage de la structure

Dans le cas d'apprentissage de la structure, tous les paramètres du réseau étant connus, nous appliquons des algorithmes capables de trouver automatiquement le graphe qui puisse représenter le mieux les dépendances (respectivement les indépendances) entre toutes les variables de notre problème. L'apprentissage de la structure reste un problème très difficile à cause de l'importance de l'espace de recherche. Cependant, si pour 2 nœuds, le nombre de structures possibles est de 3, pour 5 nœuds, ce nombre passe à 29 281. Il devient, alors, impossible de passer par tous les cas, dès que le nombre de nœuds dépasse 8. Deux grandes familles d'approches existent pour réaliser cet apprentissage. Les méthodes de recherche d'indépendance conditionnelle et les méthodes de recherche de structures à base de score. Les méthodes de la première catégorie sont issues des travaux de plusieurs équipes de recherche comme de Pearl et Verma qui ont développé les algorithmes IC et IC* [26]. Une autre équipe de chercheurs Sprites, Glymour et Scheines, a développé d'autres algorithmes comme le SGS, PC, CI et FCI [27] [28]. Les algorithmes des deux groupes sont basés sur la recherche des indépendances conditionnelles qui existent entre les variables, leurs principe passe par trois étapes que sont la construction d'un graphe non dirigé contenant les relations entre les variables par ajout d'arêtes [29] ou par suppression d'arêtes (SGS) [30], la détection des V-structures et la propagation des orientations de certains arcs. Les deux premières étapes utilisent des tests statistiques d'indépendance conditionnelle. La caractéristique principale de ces algorithmes est la détermination des relations d'indépendance entre deux variables quelconques, conditionnellement à un ensemble de variables dans le but de réduire l'espace de recherche de la meilleure structure. L'inconvénient laissant cette méthode moins utilisée que la méthode des scores est le nombre de tests à effectuer qui augmente exponentiellement avec le nombre de nœuds. Quant à la recherche de structures à base de score, elle est basée sur l'exploration heuristique de l'espace des modèles, en utilisant, des mesures de qualité qui sont les scores. D'après cette approche, il existe des structures équivalentes dites

de Markov qui forment la même fonction score, ces structures ont la particularité de représenter les mêmes indépendances conditionnelles, ce qui permet d'avoir un voisinage plus simple et un espace de recherche plus petit. Ces algorithmes peuvent procéder de deux façons différentes à savoir chercher la structure qui maximise un score (AIC, BIC, BD,...) ou encore chercher les meilleures structures et combiner leurs résultats. Une condition importante pour que ces méthodes soient applicables pratiquement, est que le score globale doit être calculé à partir des scores locaux en chaque nœud du réseau. La recherche de la meilleure structure est, cependant, non exhaustive sur l'espace de recherche à cause du nombre important de structures possibles à partir d'un nombre de nœuds. Des algorithmes comme le MWST limitent leur espace de recherche sur un espace d'arbres, l'algorithme K2 choisi, arbitrairement, un nœud racine et dirige chaque arête à partir de ce nœud, quant à l'algorithme greedy search effectue une recherche sur tous l'espace à l'aide d'opérateurs classiques : ajout d'arcs, suppression d'arcs ou inversion d'arcs, à condition que le graphe obtenu soit toujours un graphe dirigé sans cycles. Ces méthodes permettent de trouver la meilleure structure de réseaux bayésiens à adopter plus efficacement que la première méthode [22].

2.5. L'inférence

Si par définition l'apprentissage est le processus permettant la construction d'un modèle, l'inférence est l'opération qui permet de passer de la proposition à la conclusion à partir de ce modèle, c'est-à-dire, classer un nouvel exemple. Une fois la structure du graphe et la distribution des probabilités de l'ensemble des variables $\{Y_i, X\}$ sont déterminées soit avec l'aide d'un expert ou par apprentissage à partir de bases d'exemples, vient l'étape de l'inférence, consistant à calculer la probabilité conditionnelle *a posteriori* $P(Y_i/X)$ qui représente la probabilité de la variable que l'on cherche à déterminer Y_i sachant qu'on a observé l'ensemble des caractéristiques X . Notons que le problème d'inférence est, en général, NP-difficile. Pour cette raison, deux types d'algorithmes ont été développés, les méthodes d'inférence exactes et les méthodes d'inférence approchées [22].

2.5.1. Les méthodes d'inférence exactes

La méthode des « *Messages locaux* » est la première méthode d'inférence utilisée par Pearl en 1983 [31] et par Kim et Pearl en 1985. Cette méthode consiste à mettre à jour, régulièrement, les lois de probabilités, par envoi de messages de chaque variable à ses voisines dans le graphe, jusqu'à l'obtention d'un certain équilibre dans le réseau. Cet algorithme ne fonctionne que dans le cas des arbres et poly-arbres (structures arborescentes) (Fig.2.5). Si V est l'ensemble de variables ou l'information sur l'ensemble de variables du graphe, avec $V = V_1 \cup V_2$ où V_1 est l'information correspondant à l'ensemble des parents de X et V_2 est l'information correspondant à l'ensemble des enfants de X .

Deux types de messages circulent dans le graphe, λ et π et qui servent à calculer la probabilité $P(X/V = v)$:

$$P(X/V = v) = P(X/V_1, V_2) \propto P(V_2/X, V_1)P(X/V_1) \quad (2.19)$$

Comme V_1 et V_2 sont conditionnellement indépendants à X , on a $P(V_2/X, V_1) = P(V_2/X)$, ce qui nous permet d'écrire l'expression (2.20) ci-après:

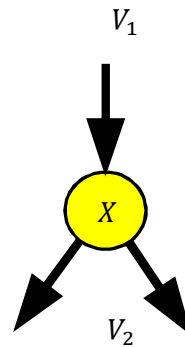


Fig.2.5 Propagation par messages locaux dans un graphe

$$P(X/V = v) \propto \lambda(X)\pi(X) \quad (2.20)$$

où $\lambda(X) = P(V_2/X)$ et $\pi(X) = P(X/V_1)$.

Ainsi, la probabilité $\lambda(X) = P(V_2/X)$ intervient comme probabilité *a posteriori* et la probabilité $\pi(X) = P(X/V_1)$ comme une vraisemblance.

Pour calculer les deux quantités $\lambda(X)$ et $\pi(X)$, il faut faire intervenir tous les enfants des enfants de X ainsi que leurs parents et tous les parents des parents de X ainsi que leurs enfants. La question qui se pose est,

comment généraliser cet algorithme pour qu'il soit applicable sur tout type de réseaux bayésiens ? Pour ce faire, Pearl en 1986 a introduit dans [32] l'algorithme « *l'ensemble de coupe* » dont le principe est de trouver un ensemble de variables appelées ensemble de coupe et d'en éliminer tous les arcs qui en sont issus afin d'obtenir un graphe sous forme d'arbre. Dans le graphe obtenu, l'inférence est effectuée en utilisant la méthode des messages locaux. Cette méthode procède à une modification topologique du graphe puis applique l'algorithme précédent. La méthode « *Arbre de jonction* » quant à elle, a été introduite par Lauritzen et Spiegelhalter en 1988 [33], Jensen, Lauritzen et Olesen en 1990 [34]. Elle est applicable sur tout type de structure du graphe, c'est-à-dire, chaînes, arbres, poly-arbres ou réseau avec boucles car elle opère des modifications importantes sur les graphes (moralisation et triangulation) afin d'obtenir l'arbre de jonction et appliquer l'algorithme des messages locaux sur les différents nœuds du réseau où chaque nœud est une clique. Comme résumé, le principe de cette méthode consiste à fusionner des nœuds pour obtenir un arbre de jonction et de factoriser la loi de probabilité jointe. L'exécution de cette méthode passe par cinq étapes en partant du réseau simple de la Fig.2.6 à savoir, la moralisation du graphe, la triangulation du graphe moral, la construction de l'arbre de jonction, l'application de l'inférence en utilisant la méthode des messages locaux et la transformation des potentiels de cliques en probabilités conditionnelles.

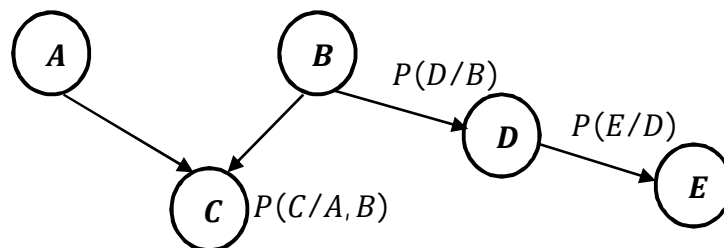


Fig.2.6 Réseau Bayésien à cinq nœuds

- Moralisation du graphe (mariage des parents)

Il s'agit de relier les parents des nœuds par des arêtes supplémentaires et supprimer les orientations de tous les arcs. Ainsi, nous obtenons un graphe dont les nœuds sont reliés uniquement par des arêtes au lieu des flèches (fig.2.7).

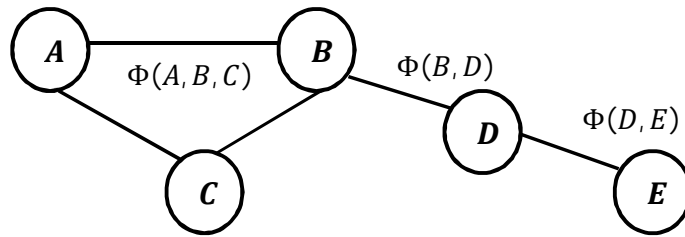


Fig.2.7 Graphe non dirigé dit moralisé

Par définition, la moralisation est une réécriture de la loi de probabilités jointe sous forme de produit des potentiels de cliques Φ_c à partir de laquelle nous pouvons, toujours, retrouver la loi initiale.

$$X = \{A, B, C, D, E\} \text{ alors } P(X) = \prod_c \Phi_c(X) \quad (2.21)$$

- Triangulation du graphe moral

Cette étape consiste à créer des sous graphes pour chaque nœud et ses parents en ajoutant des arêtes créant des raccourcis dans chaque cycle dont la longueur est d'au moins 4, pour obtenir, à la fin, un graphe moral dit triangulé, composé de cliques où chaque variable peut apparaître dans plusieurs cliques différentes (Fig.2.8).

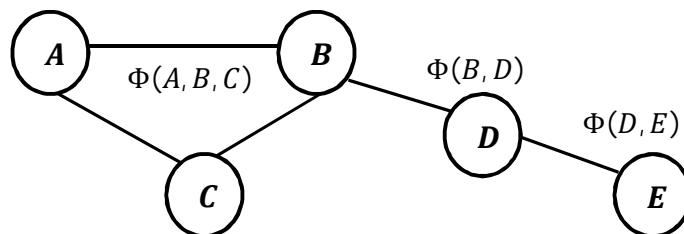


Fig.2.8 Graphe moral triangulé

La structure finale n'est pas unique. Ainsi, pour choisir la structure optimale, il faut que les nœuds d'une clique doivent appartenir au réseau bayésien sans pouvoir ajouter aucun nœud.

- Construction de l'arbre de jonction

D'après le théorème de Jensen, Lauritzen et Olesen, il est toujours possible, de construire un arbre de jonction à partir d'un graphe moral triangulé, où chaque nœud doit apparaître, au moins, une fois, dans une clique (Fig.2.9).



Fig.2.9 Arbre de jonction

Les variables de l'arbre de jonction sont des groupes de variables du graphe initial. Le chemin entre deux cliques doit absolument contenir les nœuds d'intersection. C'est la propriété de l'intersection courante.

Si l'on prend par exemple $C_1 = \{A, B, C\}$ et $C_2 = \{B, D\}$ alors $C_1 \cap C_2 = \{B\}$

Cette propriété nous mène vers la construction de l'arbre de jonction, car pour que l'information sur une variable remonte jusqu'à la clique initiale qui l'a contient, cette variable doit appartenir à toutes les cliques intermédiaires.

- Application de l'inférence en utilisant la méthode des messages locaux
L'inférence se fait sur les potentiels de cliques Φ .
- Transformation des potentiels de cliques en probabilités conditionnelles

Les lois de probabilités pour les variables d'origine peuvent être déduites à partir des potentiels de cliques Φ mis à jour, en utilisant la propriété citée précédemment.

$$\text{Si } X = \{A, B, C, D, E\} \text{ alors } P(X) = \Phi(A, B, C)\Phi(B, D)\Phi(D, E) \quad (2.22)$$

Une autre méthode adaptée pour être appliquée sur tous types de réseaux est la méthode d'élimination de variables décrite dans [35], généralisée dans [34] et qui procède à l'élimination des variables une par une avec un ordre donné. Le nombre de calculs effectué pendant l'exécution de l'algorithme dépend justement de l'ordre choisi. Cette méthode est efficace si le meilleur ordre est connu et si le réseau initial est peu dense mais avec un grand nombre de circuits.

2.5.2. Les méthodes d'inférence approchées

Les méthodes d'inférence approchées sont généralement utilisées dans le cas des réseaux bayésiens de structures complexes possédant un grand nombre de cycles ou encore des informations incomplètes pour lesquels les calculs risquent de prendre un temps important. Ainsi, on peut les utiliser pour économiser le temps des calculs. Ces méthodes sont

principalement de deux types: les méthodes exactes appliquées sur des réseaux simplifiés et les méthodes de simulation stochastique.

Dans la première catégorie, on retrouve l'approche utilisée dans [37] et [38] qui applique la méthode de propagation après simplification du réseau en mettant à zéro les probabilités trop faibles concernant les variables plus ou moins reliées l'une à l'autre. Le principe de cette approche est d'éliminer ces arcs afin d'accélérer les calculs. Cependant, il existe une deuxième approche peu différente par rapport à la première, utilisée dans [39] et [40] qui n'effectue pas la propagation des informations que sur une partie du graphe d'origine, c'est-à-dire, que sur les variables possédant les probabilités les plus élevées.

Cependant, l'idée de départ de la deuxième catégorie des méthodes d'inférence approchée était l'utilisation des connaissances *a priori* sur l'élément étudié afin de générer automatiquement et d'une façon aléatoire (stochastique) des exemples d'une base de données à partir de laquelle plusieurs estimateurs des paramètres d'une loi de probabilités pourront être calculés comme la moyenne et l'écart-type. Ces méthodes permettent de générer, à volonté, des données simulées. C'est pour cette raison qu'on retrouve le mot « simulation » dans leur appellation. En partant de ce même principe, plusieurs méthodes ont été développées comme la méthode de Monte Carlo par Chaîne de Markov dans [41], la méthode de Metropolis [42] et l'échantillonneur de Gibbs [43]. Les réseaux bayésiens sont alors des outils de simulation qui permettent à l'expert d'observer le comportement de son système dans des contextes qu'il n'est pas forcément capable de tester lui-même.

2.6. Réseaux Bayésiens particuliers

Parmi les réseaux bayésiens particuliers, nous trouvons les réseaux bayésiens dynamiques, les diagrammes d'influence qui sont des modèles étendus des réseaux bayésiens et les classificateurs bayésiens.

2.6.1. Réseau bayésien dynamique

Un réseau bayésien dynamique est une extension des réseaux bayésiens statiques qui permet de représenter l'évolution des variables dans le temps. D'après [44] et [45] il est défini par un couple de deux réseaux (R_0, R_1) où R_0 le réseau représentant les conditions initiales de la distribution de probabilités des variables à l'état initial t_0 et R_1 , le réseau à deux instants différents T et $T - 1$ décrivant la transition des variables du réseau de $t - 1$ à t . La probabilité jointe pour l'ensemble des variables V du réseau pour ces deux instants est :

$$P(V_T/V_{T-1}) = \prod_{t=T-1}^T \prod_{x \in V} P(x(t)/pa(x(t))) \quad (2.23)$$

Cette représentation permet de modéliser un réseau bayésien virtuellement infini qui en est le développement dans le temps d'un réseau initial à partir d'un instant 0. La structure de ce réseau dynamique n'est qu'un ensemble de réseaux statiques reliés par des arcs temporels (Fig.2.10) où chaque réseau est une partie du réseau global à un instant donné. Parmi les réseaux bayésiens dynamiques, les plus connus et les plus utilisés, le filtre de Kalman qui est un modèle estimant les états d'un système dynamique à partir d'une série de mesures incomplètes ou bruitées [46].

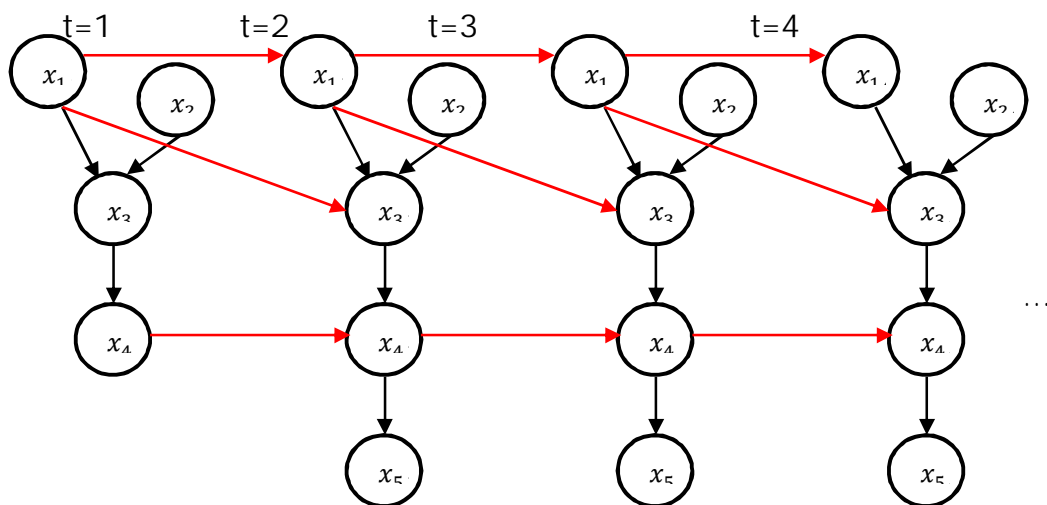


Fig.2.10 Réseau bayésien dynamique

2.6.2. Diagrammes d'influence

Il s'agit de modèles étendus des réseaux bayésiens modélisant une situation de décision. Un diagramme d'influence est un graphe dirigé acyclique contenant trois types de nœuds, à savoir des nœuds de décision (action), des nœuds de chance ou de probabilités des variables et des nœuds d'utilités qui représentent la qualité, l'ordre de préférence ou le coût accordée à chaque décision prise. Les arcs vers les nœuds de décision indiquent les informations connues avant que la décision ne soit prise et tous les autres arcs indiquent les dépendances probabilistes. La procédure de calculs est similaire à celle des réseaux bayésiens en passant par : la moralisation du réseau de valuation dans lequel ont été supprimés tous les arcs entrant dans les nœuds de décision, la suppression des nœuds d'utilité, la triangulation du réseau obtenu et la création de l'arbre de jonction dans lequel les calculs seront effectués [47] en utilisant l'algorithme d'inférence « arbre de jonction » (Fig.2.11).

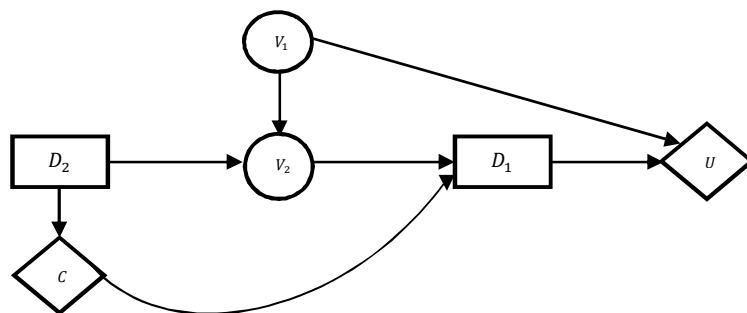


Fig.2.11 Exemple d'un diagramme d'influence

2.6.3. Classificateurs bayésiens

Dans le cas d'un problème de classification, la structure du réseau est simplifiée par une hypothèse dite l'hypothèse naïve de Bayes qui considère que les caractéristiques x_i sont deux à deux indépendantes conditionnellement à la variable classe C_j . Pour ce genre de réseaux, on retrouve deux niveaux de nœuds, le premier niveau est le niveau du nœud parent « classe » considéré comme quantité non observée et qu'il faut déterminer. Le deuxième est celui des nœuds enfants « caractéristiques », considérés comme étant des variables observées caractérisant l'élément étudié (Fig.2.12).

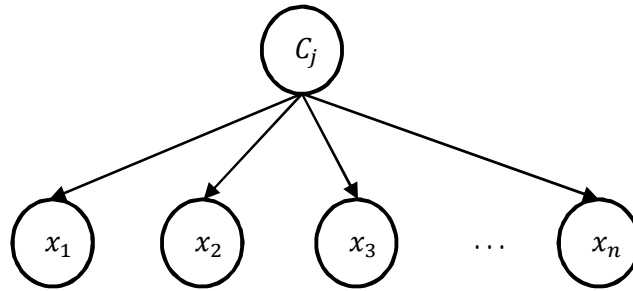


Fig.2.12 Réseau bayésien naïf

Ainsi, le calcul de la probabilité jointe devient plus simple.

$$P(C_j, x_1, x_2, \dots, x_n) = P(C_j) \times P(x_1, x_2, \dots, x_n / C_j) = P(x_1, x_2, \dots, x_n) \times P(C_j / x_1, x_2, \dots, x_n) \quad (2.24)$$

D'ici, on obtient la probabilité *a posteriori* de la classe sachant les caractéristiques donnée par la relation suivante :

$$P(C_j / x_1, x_2, \dots, x_n) = \frac{P(C_j) \times P(x_1, x_2, \dots, x_n / C_j)}{P(x_1, x_2, \dots, x_n)} = \frac{P(C_j) \times P(x_1 / C_j) \times P(x_2 / C_j) \times \dots \times P(x_n / C_j)}{P(x_1, x_2, \dots, x_n)} \quad (2.25)$$

Selon certains chercheurs, cette configuration peut dans certains cas se montrer insuffisante dans le sens où, pratiquement, il existe toujours des relations entre les caractéristiques [48] et [49] bien que cette dernière donne de très bons résultats de classification [50]. Parmi les algorithmes qui prennent en considération l'existence de cette relation, le TANB (Tree Augmented Naive Bayes) qui est un réseau bayésien naïf augmenté par un arbre dont la structure entre la variable « classe » et les « caractéristiques » est la même que celle du réseau bayésien naïf en rajoutant que chaque caractéristique est reliée à toutes les autres (Fig.2.13), le but de cette configuration est l'augmentation du pouvoir de classification en recherchant le meilleur arbre.

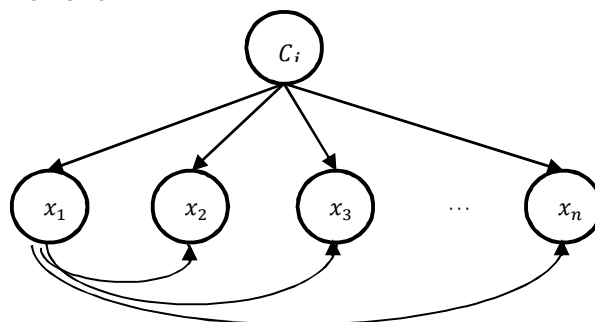


Fig.2.13 Réseau bayésien augmenté par un arbre

Dans [51], il est proposé d'augmenter le réseau naïf par une forêt en utilisant l'algorithme FANB (Forest Augmented Naive Bayes) dont le principe est de trouver l'ensemble optimal d'arbres disjoints sur l'ensemble des variables caractéristiques. Cependant, il existe une autre approche plus récente appelée les multi-nets [52][53] qui a apporté deux points importants dont le premier est que les relations de causalité et d'indépendances entre les attributs ne sont pas forcément les mêmes selon les différentes modalités du nœud principal « classe » et que la structure représentant les liaisons entre les attributs pour une classe fixée est beaucoup plus simple que de la structure qui relie toutes les variables (Fig.2.14).

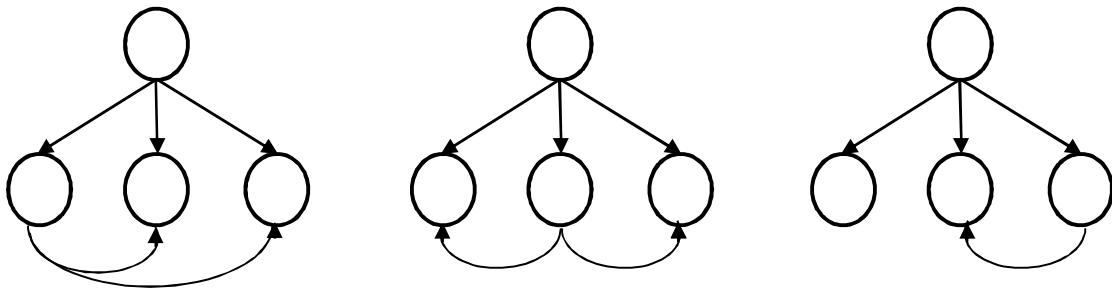


Fig.2.14 Réseau bayésien multi-nets

2.7. Conclusion

Dans ce chapitre, nous avons présenté les réseaux bayésiens et nous avons passé en revue les différents algorithmes d'apprentissage ainsi que les méthodes d'inférence les plus connues. Par suite, nous avons cité quelques modèles particuliers des réseaux bayésiens parmi lesquels, les classificateurs bayésiens naïfs que nous utiliserons dans notre application qui consiste à réaliser le système de contrôle automatique et de prise de décision de la serre agricole. La première étape de conception de ce système de classification est la construction du modèle qui consiste à définir les paramètres qui dans plusieurs cas s'ils ne sont pas donnés par les experts, doivent être calculés à partir d'une base de données recueillies à partir des différents capteurs grâce aux différents algorithmes d'apprentissage qui existent.

Chapitre 3

Application des RB à une serre agricole

3.1. Introduction

Après avoir introduit la théorie des réseaux bayésiens dans le chapitre précédent, nous allons procéder aux relevés de mesures des paramètres du microclimat d'une serre agricole dans laquelle sont placés différents types de capteurs pour, ensuite, décider si les meilleures conditions qui conviennent à la croissance d'une plante donnée sont présentes. Dans notre application, nous nous intéressons au cas de l'olivier dont les paramètres du microclimat qui lui conviennent le mieux nous ont été fournis par un expert agricole de la région d'Alger. Notons que l'approche multi-capteurs que nous avons adoptée peut être vue comme un problème de reconnaissance de formes et, plus précisément, de classification automatique où les classes constituent les différents états du climat de la serre et les formes représentent l'ensemble des observations ou mesures multi-capteurs relevées. Ainsi, dans la prochaine section, nous allons présenter l'effet des paramètres du microclimat sur la croissance des plantes. Dans la troisième section, il

s'agit des réseaux multi-capteurs utilisés dans une serre agricole. Dans la quatrième section, nous présentons notre application concernant le microclimat lié à l'olivier. Les résultats seront présentés dans la cinquième section et nous terminons par une conclusion.

3.2. Paramètres du microclimat dans une serre

Les différents paramètres climatiques qui sont, généralement, pris en compte dans un microclimat d'une serre agricole sont, les températures, ambiantes et du sol, l'humidité et le rayonnement [54]. Ces informations issues des différents capteurs sont transmises vers une station de contrôle qui assure l'acquisition, le traitement et l'analyse des données. Notons que, généralement, le microclimat désigne, les conditions climatiques limitées à une région géographique restreinte, significativement distincte du climat général de la zone où se situe cette région [55]. Dans notre application, ce terme désigne le climat interne de la serre agricole qui a une influence directe sur le développement des différentes parties de la plante. C'est ainsi que tout l'intérêt est porté sur la maîtrise et le contrôle des paramètres environnementaux dans la serre en vue d'obtenir une récolte à rendement élevé. Notons que, en plus de l'avantage du contrôle du microclimat, la végétation dans une serre est protégée contre les animaux nuisibles, les variations météorologiques brusques comme la chaleur ou froid excessifs, les tempêtes, la sécheresse, le vent etc. De plus, la serre favorise, la croissance des plantes non adaptées à une région donnée ainsi que celle des plantes hors de leurs saisons. Pour arriver à satisfaire tous ces avantages, il est nécessaire de contrôler les paramètres qui présentent une influence sur le microclimat à l'intérieur de la serre. Parmi ces paramètres, les plus influents sont certainement, la température, l'humidité et le rayonnement.

3.2.1. Influence de la température

En biologie, de nombreuses réactions chimiques dépendent de l'évolution de la température. Par exemple, lorsque celle-ci est augmentée de 10°C, la vitesse de la réaction double. A l'inverse, la réaction devient plus lente et se produit en deux fois plus de temps. Cet effet peut être observé quand la température des racines n'est pas la même que celle des feuilles. Comme exemple, prenons le cas d'une plante placée sur le sol. Alors que son feuillage exposé au soleil atteint une température de 22° C, ses racines sont soumises à une température de 10°C. Les feuilles travaillent donc deux fois plus vite que les racines. Le résultat observé est la brûlure des feuilles du fait que les racines dont le fonctionnement est au ralenti ne parviennent pas à irriguer la partie aérienne [56]. Ainsi, la température représente l'un des paramètres les plus importants pour la gestion du climat. Il est donc primordial de contrôler la température ambiante et du sol à l'intérieur de la serre, soit naturellement en commandant des toits ouvrant soit, en introduisant des chauffages et des refroidisseurs artificiels.

3.2.2. Influence de l'humidité

L'évaporation est définie comme étant l'émission d'eau à l'état de vapeur par les feuilles des végétaux dans l'atmosphère. C'est un mécanisme essentiel qui permet le maintien de l'équilibre hydrique des végétaux et la régulation de la température de l'arbre. Seul 10 % environ de l'eau puisée dans le sol sert à la photosynthèse, le reste s'en va lors de l'évaporation. Ainsi, une transpiration efficace est favorisée par un air ambiant sec et en mouvement. Quand l'humidité relative est plus faible à l'extérieur de la feuille, l'eau de la plante peut facilement s'évaporer. Aussi, plus l'air est sec, plus il y a d'échanges gazeux entre la feuille et son environnement. Dans de telles conditions, la plante absorbe bien l'eau du sol avec ses nutriments. Quand l'air est immobile, l'eau évaporée par les feuilles s'accumule autour de celles-ci. La plante transpire alors de

moins en moins et son feuillage se réchauffe, chose qui risque de déséquilibrer le fonctionnement normal de la croissance de la plante. Il est donc vital d'adapter le degré d'humidité d'air aux besoins de la plante par la mise en place de ventilateurs et d'humidificateurs.

3.2.3. Influence de la lumière

La photosynthèse et la respiration sont deux phénomènes que produisent les plantes vertes. Sous une lumière particulièrement intense, la plante réalise la photosynthèse pendant laquelle elle produit des nutriments et de l'oxygène. Pendant l'absence de la lumière, la photosynthèse cesse et la respiration devient le processus dominant pendant lequel la plante consomme les nutriments produits pour sa croissance. Lorsque l'intensité de la lumière augmente au-delà d'un certain seuil, la chaleur produite augmente, ce qui endommage les cellules délicates des plantes tout en augmentant la vitesse de transpiration. Ceci aurait, pour effet, le dessèchement des feuilles. Ainsi, pour optimiser les conditions de croissance des plantes, la durée et l'intensité de la lumière peuvent être contrôlées en utilisant une source de lumière artificielle.

Pour pouvoir mesurer et contrôler l'ensemble de ces paramètres, des réseaux de capteurs de différents types sont installés à l'intérieur de la serre.

3.3. Réseaux multi-capteurs dans une serre agricole

Les serres sont des environnements de mesures exigeants et qui représentent un nombre important de contraintes comme le risque de condensation, le réchauffement causé par le rayonnement solaire, l'augmentation de l'humidité interne et la poussière. Ainsi, les instruments de mesures utilisés doivent être bien choisis pour fonctionner correctement dans de telles conditions. Les points importants à prendre en compte sont, entre autre, la précision, la stabilité à long

terme, le temps de réponse ainsi que la facilité d'intégration de ces capteurs dans les installations de contrôle et de commande sous forme d'un réseau. En effet, un réseau multi-capteurs est un ensemble de capteurs capables de récolter et de transmettre vers une station (ordinateur) des données environnementales d'une manière autonome. Dans la technologie sans fil, on parle de micro-capteurs dont chacun est représenté par un nœud non obligatoirement prédéterminée. Cependant, ils peuvent être dispersés, aléatoirement dans une zone géographique appelée *champ de captage* correspondant au terrain d'intérêt pour le phénomène capté [57]. Les données collectées par les différents capteurs sont envoyées vers un nœud central appelé *Point de collecte* et elles seront transmises vers la station de l'utilisateur via internet, satellite ou autre système de transmission (Fig.3.1).

Dans la serre, le capteur doit avoir un emplacement bien représentatif du paramètre que l'on veut mesurer. Quelques recommandations ont été données pour bien positionner le capteur selon sa nature. En effet, le capteur de température doit être placé dans le même espace que la plante. Un capteur installé près de la toiture, devant la porte d'entrée ou encore à côté des conduites de chauffage n'est pas représentatif de la température ambiante de la plante. Le capteur d'humidité ne doit jamais être placé à proximité des conduites de chauffage ni d'humidificateurs par exemple alors que le capteur de luminosité doit être disposé de façon à absorber un maximum de rayonnements. Toutes ces informations contribuent à la gestion du chauffage ou de refroidissement, de l'humidification de l'air interne et de l'éclairage de la serre.

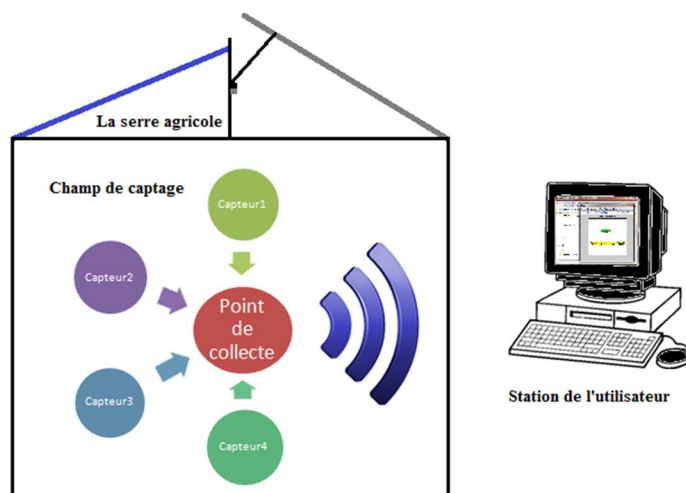


Fig.3.1 Multi-capteurs dans une serre agricole

3.4. Méthodes de contrôle des serres agricoles

Lorsqu'on parle de contrôle automatique ou intelligent du microclimat, il s'agit d'évaluer l'état de ce dernier à chaque instant grâce aux informations issues du réseau multi-capteurs. Pour ce faire, on a recours à plusieurs méthodes de la plus classique aux évoluées.

3.4.1. Méthode classique

Dans les serres traditionnelles, aucune option de contrôle dynamique du climat n'était faisable. Grâce aux progrès de la science, aujourd'hui, il existe plusieurs techniques et dispositifs capables de gérer le microclimat de manière manuelle ou automatique. Pour commander les différents paramètres climatiques de la serre selon la plante, l'opérateur doit relever les différentes mesures des capteurs disposés dans les serres à des instants réguliers afin d'agir, par la suite, sur les différents actionneurs de type ventilateurs, chauffage, éclairage et humidificateurs en vue de modifier le paramètre désiré. Cette méthode nécessite le déplacement permanent de l'opérateur dans la serre et le temps de réponse est relativement long. En effet, on ne modifie le paramètre déréglé qu'une fois l'événement apparu, vérifié et analysé, chose pouvant entraîner de graves conséquences sur la quantité et la qualité des

récoltes, d'où l'intérêt de disposer d'un contrôle automatique du microclimat.

3.4.2. Méthodes modernes

Les systèmes de surveillance automatiques permettent de détecter les défaillances à des instants très rapprochés. Deux catégories de méthodes sont appliquées à ce type de contrôle, à savoir, les méthodes avec modèles et les méthodes sans modèles.

Les approches de contrôle avec modèles ont pour objectif de comparer les mesures effectuées sur le système réel avec les informations fournies par le modèle. Ainsi, tout écart entre les mesures et les sorties du modèle, est synonyme d'anomalies. Les outils, issus de la théorie de décision, sont, alors, utilisés pour déterminer exactement le type d'anomalie qui peut être dû, soit à une perturbation externe, une défaillance d'un capteur de mesure ou finalement d'un état normal de fonctionnement. Le modèle est donné sous forme d'équations algébriques ou différentielles, déduit à partir des lois de la physique ou par identification, mais avant toute utilisation, chaque modèle doit obligatoirement être validé expérimentalement. Ces méthodes peuvent se décomposer en deux catégories, les méthodes de redondance physique et analytique et les méthodes d'estimation paramétrique. Pour les méthodes de redondance physique, il s'agit de doubler ou de tripler les capteurs du système pour s'assurer de la validité des mesures et pour distinguer s'il s'agit d'une défaillance d'un capteur ou d'un dérèglement du système lui-même et qu'il faut corriger [58]. Si ces composantes placées dans le même environnement fournissent des signaux identiques, on considère alors qu'ils sont dans un fonctionnement nominal. Dans le cas contraire, on estime qu'une défaillance s'est produite dans au moins l'un des capteurs. Cette méthode a pour but d'augmenter la fiabilité des mesures. Parmi les caractéristiques de la redondance matérielle, la simplicité de la conception, le coût élevé de la mise en œuvre dans le cas ou plusieurs

paramètres doivent être mesurés, ce qui conduit à des installations plus ou moins encombrantes. Quant aux méthodes de redondance analytique, la solution ne peut être mise en œuvre que si l'on dispose d'un modèle mathématique précis du système. Cette méthode consiste à utiliser des informations supplémentaires issues des modèle et celles provenant des capteurs [59]. Le but est d'estimer l'état du système afin de le comparer à son état réel. Pour les méthodes d'estimation paramétrique, le principe consiste à estimer en continu les paramètres du système en utilisant les mesures d'entrée-sortie, afin de les comparer aux paramètres de référence de l'état normal du procédé. Pour détecter l'existence d'une anomalie, il faut comparer en permanence les paramètres estimés et ceux théoriques calculés à partir du modèle.

Dans le cas de méthodes sans modèles, l'élaboration du modèle des systèmes naturels ou industriels, le mode d'évolution peut être dynamique, aléatoire et non linéaire. Ce modèle est complexe voire même impossible dans certains cas. Pour les applications impliquant ce type de systèmes, les seules méthodes de contrôle opérationnelles sont celles n'utilisant pas de modèles formels de fonctionnement. Dans cette catégorie, on trouve les outils statistiques de traitement de signal et toutes les méthodes basées sur l'intelligence artificielle telles que la reconnaissance de formes [60]. Les techniques basées sur les outils statistiques de traitement de signal supposent que les différents signaux en provenance de la chaîne de mesure possèdent certaines propriétés statistiques, moyenne et variance, alors pour contrôler et vérifier si le système est dans son fonctionnement normal ou désiré, il faut comparer, régulièrement, ces caractéristiques identifiées avec ceux déterminés, préalablement, sur un échantillon de mesures lors du fonctionnement nominal du système en utilisant des tests. Une grande variété de tests existe et parmi les plus utilisés, on peut citer le test de franchissement de seuils et le test de moyenne et test de variance. Le principe du test de franchissement de seuils est de comparer régulièrement les mesures

issues des capteurs avec des seuils prédéterminés par des spécialistes du domaine concerné. Dans le cas où un seuil est franchi, le système de contrôle déclenchera une alarme, voire, affichera un message, à titre informatif pour l'opérateur et enverra un signal pour le système de commande responsable de la correction de ce dérèglement. Parmi les caractéristiques de cette méthode, la simplicité de la mise en œuvre qui reste, cependant, une technique sensible aux fausses alarmes et ne permet pas l'établissement d'un diagnostic du type de l'anomalie. La méthode de test de la moyenne propose de comparer la moyenne d'un certain nombre de mesures avec la moyenne prédéterminée du fonctionnement nominal, chose qui permet, aussi, de vérifier, si, le système est dans un état normal ou non. Le test de la variance, quant à lui, a pour principe de calculer la variance du signal de mesures, tant que cette valeur se situe autour de la valeur nominale préfixée, ce qui signifie que le fonctionnement peut être considéré normal. Par ailleurs, réaliser le contrôle d'un système par reconnaissance de formes consiste à décrire le système à un instant donné par n caractéristiques qui constituent le vecteur forme et construire une fonction de décision permettant de classer cette forme dans une classe parmi d classes possibles en minimisant un critère donné. Les classes correspondent aux différents états que l'on cherche à reconnaître. Le système à un instant donné est ainsi représenté par un point dans un espace de représentation à n paramètres. La fonction de décision partitionne l'espace de représentation en d régions, chaque région correspondant à un état donné.

Un ensemble de capteurs permet l'enregistrement de signaux sur le système qui sont ensuite filtrés et mis en forme. Les caractéristiques sont ensuite extraites des signaux et forment les entrées de la fonction de décision qui classe la forme observée à l'instant t dans une des classes ou états possibles. Construire un système de contrôle par reconnaissance de formes consiste donc, dans un premier temps, à choisir les caractéristiques du vecteur forme puis à apprendre la fonction de décision

à partir d'une base de données d'apprentissage en définissant les frontières des différentes régions. La base d'apprentissage doit être construite avec soin car les performances des fonctions de décision dépendent de la qualité des exemples utilisés pour l'apprentissage. L'ensemble des états que l'on cherche à reconnaître doit être représenté dans la base par un nombre d'exemples suffisants et correctement expertisés. Les exemples doivent correspondre à des enregistrements effectués pendant des périodes de temps où le système est stationnaire dans un état en évitant, si possible, les périodes de transitions d'un état vers un autre, au cours desquelles l'état du système est indéterminé. Les informations enregistrées sur le système via les capteurs sont des signaux. L'extraction de caractéristiques consiste à extraire des signaux préalablement filtrés, les informations pertinentes pour reconnaître les états que l'on recherche. Un signal enregistré au cours du temps est une dimension temporelle. Or, le classement en différentes classes par la fonction de décision est une opération statique où le temps n'intervient pas. Le système est représenté à un instant t par n caractéristiques auxquels on associe une classe. Notons que les caractéristiques des exemples de la base d'apprentissage doivent être extraites avec beaucoup de soins, la présence de valeurs erronées pouvant dégrader les performances en déformant les classes et donc introduire une erreur sur les frontières de décision calculées lors de la phase d'apprentissage. La règle de décision D est apprise à l'aide de la base d'apprentissage qui est composée d'un ensemble de m individus $X = (x_1, x_2, x_3, \dots, x_m)$, avec x_j un vecteur forme composé de n caractéristiques $x_j = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{jn})$. La validation du système de contrôle consiste à vérifier les capacités du système à classer de nouveaux exemples. Ainsi, le pourcentage d'exemples mal classés par le système dans la base d'apprentissage, appelée erreur en apprentissage, ne peut pas constituer un critère de validation. Seul le pourcentage de nouveaux exemples mal classés appelé erreur en test, peut fournir un indice de performance. Il est donc

nécessaire de réserver, dans la base d'apprentissage, des exemples qui serviront ultérieurement à la validation. La base de données est divisée en deux parties : une base d'apprentissage permettant l'apprentissage des règles de décision et une base de test permettant l'estimation de l'erreur réelle.

3.5. Application des réseaux bayésiens à la serre

Il s'agit dans cette partie de travail, d'appliquer les réseaux bayésiens dans un contexte de fusion de données pour surveiller un microclimat pour la culture de l'olivier à l'intérieur d'une serre agricole.

3.5.1. Création de la base de données simulée

Pour un système de classification basé sur l'apprentissage, la composante la plus importante est la base de données qui permettra la construction du modèle de classification afin de distinguer les différents états du microclimat. Dans cette base de données, chaque vecteur descripteur X doit avoir sa classe d'appartenance C_i . Les données réelles n'étant pas disponibles, nous avons constitué une base de données simulée composée d'un certain nombre de cas fictifs de microclimat en se basant sur les valeurs optimales des caractéristiques climatiques pour un olivier sous serre fournies par un expert de la région d'Alger et qui sont résumées dans le tableau 3.1.

Tableau 3.1 Valeurs optimales des caractéristiques climatiques pour un olivier sous serre.

Propriétés	Température ambiante T_a (°C)	Température du sol T_s (°C)	Humidité ambiante H_a (%)	Rayonnement R_a (Lux)
Valeurs des paramètres	$22 \leq T_a \leq 24$	$18 \leq T_s \leq 22$	$90 \leq H_a \leq 100$	$100 \leq R_a \leq 200$

La structure de cette base respecte les paramètres thermiques, hygrométriques et lumineux connus du microclimat adéquat pour un olivier sous serre et qui seront fournis par le programme établi.

L'avantage d'avoir une base de données simulée est de pouvoir intégrer tous les cas possibles sans aucune difficulté, ce qui permet d'avoir une base la plus complète possible. Pour cela, nous avons organisé notre travail en deux fichiers. Le premier fichier appelé « *base_donnees.m* » pour la base de données qui sera, par la suite, divisée en deux ensembles, le premier pour l'entraînement du classificateur, le second pour le test et la validation du système de contrôle ainsi réalisé. Le deuxième fichier que nous avons nommé « *reseau_bayesien.m* » est celui qui contient le code de notre application. Les quatre paramètres sont générés à l'aide de l'instruction « *rand* », en respectant les limites de chacun d'eux selon son adéquation ou non pour la culture en question. Ainsi, les différentes classes correspondent à tous les cas possibles de l'état du microclimat, selon l'état de chaque paramètre. Dans le cas où le microclimat est bon, cela implique, différents cas.

Tableau 3.2 Les différents états des paramètres climatiques

Les états Les paramètres	Etat inadéquat inférieur	Etat adéquat	Etat inadéquat supérieur
La température ambiante T_a (°C)	[-20,22]	[22,24]	[24,70]
La température du sol T_s (°C)	[-20,18]	[18,24]	[22,70]
L'humidité ambiante H_a (%)	[0,90]	[90,100]	–
Le rayonnement R_a (Lux)	[0,100]	[100,200]	[200,1000]

L'avantage de connaître le paramètre et/ou les paramètres déréglés est de pouvoir intervenir efficacement et rapidement. Pour dénombrer tous les cas possibles, nous avons fait appel à l'analyse combinatoire. Comme l'ordre, dans notre cas, n'importe pas, il s'agit donc de combinaisons d'un nombre de p éléments pris parmi n pour chaque paramètre tel que :

$$C_n^p = \frac{n!}{p!(n-p)!} \quad (3.1)$$

Pour les trois paramètres T_a , T_s et R_a , nous prenons un seul état parmi trois et pour l'humidité H_a , nous prenons un état parmi 2 simultanément (Tableau 3.2).

En tous, nous avons obtenu $N = 54$ cas différents calculés de la manière suivante :

$$N = C_3^1 \times C_2^1 \times C_3^1 \times C_3^1 = 3 \times 2 \times 3 \times 3 = 54 \text{ classes (états du microclimat).}$$

Ainsi, cette règle peut être généralisée sous la forme de l'expression 3.2 suivant le nombre de capteurs et les états de chaque paramètre utilisés dans une application donnée.

$$N = \prod_n (C_k^1)^n \quad (3.2)$$

où N est le nombre de tous les cas possibles, n , le nombre de paramètres utilisé dans l'application et k , le nombre d'états de chaque paramètre.

Le tableau 3.3 résume tous ces cas. Dans ce tableau, le signe « + » indique que le paramètre est dans cet état. Le signe « - » indique que le paramètre n'est pas dans cet état. Dans la plupart du temps, comme dans notre cas, la banque de données, déjà classées, est divisée en deux ensembles, l'ensemble d'entraînement sur lequel le classificateur fait son apprentissage et l'ensemble de tests sur lequel on peut évaluer ses performances.

3.5.2. Outil de programmation

Nous avons utilisé la version 7.12.0.635 de Matlab comme logiciel de programmation doté de la boîte à outils des réseaux Bayésiens BNT (Bayesian Network Toolbox). La BNT est une bibliothèque de fonctions Matlab conçue pour la création, l'apprentissage et l'inférence de modèles graphiques, en général. Elle offre plusieurs densités de probabilités utilisées pour la modélisation des variables continues. Dans le cadre de l'apprentissage, la BNT est capable d'estimer les paramètres du réseau à partir de données complètes par des algorithmes préprogrammés comme « *le maximum de vraisemblance* » en utilisant l'instruction *learn_params*

et « *le maximum a posteriori* » en utilisant *bayes_update_params*, ou à partir de données incomplètes par la méthode *EM* en utilisant *learn_params_em*.

Tableau 3.3 Les différentes classes du microclimat

Etats Classes	Température ambiante adéquate	Température du sol adéquate	Humidité adéquate	Rayonnement adéquat	Température ambiante inadéquate supérieure	Température ambiante inadéquate inférieure	Température du sol inadéquate supérieure	Température du sol inadéquate inférieure	Humidité inadéquate inférieure	Rayonnement Inadéquat supérieur	Rayonnement Inadéquat inférieur
C ₁	+	+	+	+	-	-	-	-	-	-	-
C ₂	-	-	-	-	+	-	-	+	+	-	+
C ₃	-	-	-	-	-	+	+	-	+	-	+
C ₄	-	-	-	-	-	+	-	+	+	+	-
C ₅	-	-	-	-	+	-	+	-	+	+	-
C ₆	-	-	-	-	+	-	-	+	+	+	-
C ₇	-	-	-	-	-	+	+	-	+	+	-
C ₈	-	-	-	-	+	-	+	-	+	+	-
C ₉	-	-	-	-	-	-	-	-	-	-	-
C ₁₀	-	+	+	+	+	-	-	-	-	-	-
C ₁₁	-	+	+	+	-	+	-	-	-	-	-
C ₁₂	+	-	+	+	-	-	+	-	-	-	-
C ₁₃	+	-	+	+	-	-	-	+	-	-	-
C ₁₄	+	+	-	+	-	-	-	-	+	-	-
C ₁₅	+	+	+	-	-	-	-	-	-	+	-
C ₁₆	+	+	+	-	-	-	-	-	-	-	+
C ₁₇	-	-	+	+	+	-	+	-	-	-	-
C ₁₈	-	-	-	+	-	+	-	+	-	-	-
C ₁₉	-	-	+	+	+	-	-	+	-	-	-
C ₂₀	-	-	+	+	-	+	+	-	-	-	-
C ₂₁	-	+	-	+	+	-	-	-	+	-	-
C ₂₂	-	+	-	+	-	+	-	-	+	-	-
C ₂₃	-	+	+	-	+	-	-	-	-	+	-
C ₂₄	-	+	+	-	+	-	-	-	-	-	+
C ₂₅	-	+	+	-	-	+	-	-	-	+	-
C ₂₆	-	+	+	-	-	+	-	-	-	-	+
C ₂₇	+	-	-	+	+	-	-	-	+	-	-
C ₂₈	+	-	-	+	-	+	-	-	+	-	-

C_{29}	+	-	+	-	-	-	+	-	-	+	-
C_{30}	+	-	+	-	-	-	+	-	-	-	+
C_{31}	+	-	+	-	-	-	-	+	-	+	-
C_{32}	+	-	+	-	-	-	-	+	-	-	+
C_{33}	+	+	-	-	-	-	-	-	+	+	-
C_{34}	+	+	-	-	-	-	-	-	+	-	+
C_{35}	-	-	-	+	+	-	+	-	+	-	-
C_{36}	-	-	-	+	+	-	+	-	+	-	-
C_{37}	+	+	-	+	+	-	-	+	+	-	+
C_{38}	-	-	-	+	-	+	-	+	+	-	-
C_{39}	-	-	+	-	+	-	+	-	-	+	-
C_{40}	-	-	+	-	+	-	+	-	-	-	+
C_{41}	-	-	+	-	+	-	-	+	-	+	-
C_{42}	-	-	+	-	-	+	+	-	-	+	-
C_{43}	-	-	+	-	+	-	-	+	-	-	+
C_{44}	-	-	+	-	-	+	+	-	-	-	+
C_{45}	-	-	+	-	-	+	-	+	-	+	-
C_{46}	-	-	+	-	-	+	-	+	-	-	+
C_{47}	-	+	-	-	+	-	-	-	+	+	-
C_{48}	-	+	-	-	+	-	-	-	+	-	+
C_{49}	-	+	-	-	-	+	-	-	+	+	-
C_{50}	-	+	-	-	-	+	-	-	+	-	+
C_{51}	+	-	-	-	-	-	+	-	+	+	-
C_{52}	+	-	-	-	-	-	+	-	+	-	+
C_{53}	+	-	-	-	-	-	-	+	+	+	-
C_{54}	+	-	-	-	-	-	-	+	+	-	+

3.5.3. Création du réseau bayésien

Le réseau bayésien a été créé par l'instruction *mk_bnet*. Il est constitué de cinq nœuds. Le nœud *classe=microclimat* est reconnu à partir des autres variables représentant les caractéristiques (Fig.3.2). La structure du réseau bayésien est représentée sous sa forme la plus simple avec une indépendance totale des caractéristiques conditionnellement au nœud microclimat. Les caractéristiques choisies sont les paramètres climatiques. Ces paramètres sont considérés comme des variables continues et les éventuels états du microclimat sont représentés par la variable microclimat qui est considérée comme une variable discrète.

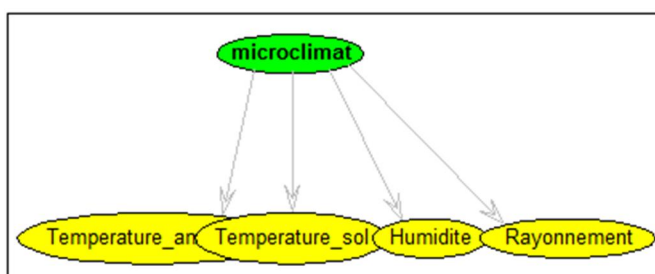


Fig.3.2 Le réseau Bayésien naïf des paramètres de la serre créée

Pour arriver à prendre une décision sur la nature du microclimat parmi les 54 cas possibles, notre algorithme passe par les quatre étapes utilisées lors de la fusion de données multi-capteurs, réalisées par le biais d'une méthode de classification basée sur les réseaux bayésiens. Notons que nous effectuons seulement l'apprentissage des paramètres car la structure est celle du réseau bayésien naïf.

3.5.3.1. Modélisation

Cette étape est fondée sur l'utilisation des probabilités conditionnelles $P(X/C_j)$ pour la représentation de toutes les variables dans un cadre commun. Le modèle proposé est un modèle où deux types de lois de probabilités sont mélangés, la loi discrète et la loi continue gaussienne. Après l'observation des deux types de variables, nous avons constaté que les différents états du microclimat sont considérés comme de simples étiquettes des classes dont les valeurs sont prises dans l'ensemble des entiers telle que $C_j \in \{1,2,3,\dots,54\}$. L'observation de l'évolution des

paramètres climatiques à partir de bases de données météorologiques provenant d'internet [61] nous a conduits à considérer que ces variables peuvent être modélisées par des densités de type gaussien. Lorsqu'on représente un échantillon de mesures de la température pendant 24H en fonction du temps (Tableau 3.4), nous obtenons un nuage de points qui a pour forme la courbe de la loi normale conformément à la figure 3.3.

Tableau 3.4 Les valeurs de températures d'une région méditerranéenne

Heure	Températures
7h	20.3°C
8h	21.8°C
9h	24°C
10h	25.6°C
11h	26.9°C
12h	28.1°C
13h	29.5°C
14h	30.4°C
15h	31.1°C
16h	31.5°C
17h	32.3°C
18h	32°C
19h	31.1°C
20h	30.4°C
21h	27.6°C
22h	26.2°C
23h	24.8°C
0h	23.9°C
1h	23.2°C
2h	22.4°C
3h	21.8°C
4h	21.1°C
5h	21.1°C
6h	20.6°C

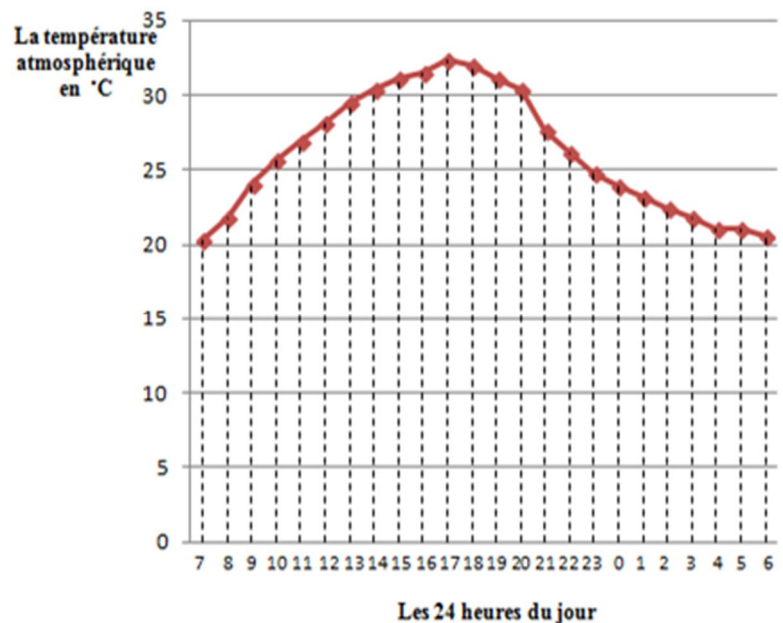


Fig.3.3 Courbe de la température climatique d'une région méditerranéenne.

3.5.3.2. Estimation des paramètres

A partir du moment où l'étape de modélisation fait intervenir des distributions de probabilités, cette étape consiste à effectuer leurs calculs. Ces paramètres sont estimés, en utilisant, les algorithmes d'apprentissage de paramètres sans données manquantes qui est implémenté dans la BNT par l'instruction « *learn_params* ». Prenons les notations suivantes pour les cinq variables du réseau bayésien (Fig.3.4).

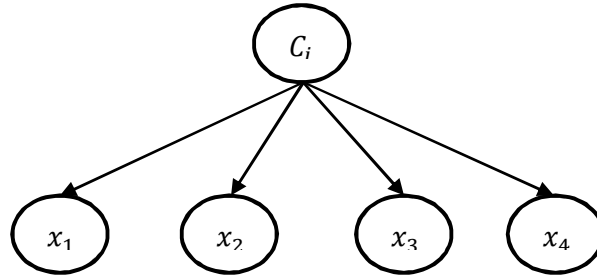


Fig.3.4 Les notations utilisées dans le réseau bayésien

Dans cette figure, $x_1 = Ta$, $x_2 = Ts$, $x_3 = Ha$, $x_4 = Ra$ et $C_j = \text{microclimat}$. Les paramètres à calculer sont de deux types, à savoir :

Le nœud C_j qui est un nœud parent et sans parents. La probabilité à calculer est $P(C_j)$. Ce terme est donné par la fréquence d'apparition de chaque classe dans la base de données puisqu'il s'agit d'une variable discrète (expression 3.3). Dans notre cas toutes les classes sont équiprobables.

$$P(C_1) = P(C_2) = \dots = P(C_{54}) = \frac{N(C_j)}{N} = \frac{N(C_j)}{\sum_{j=1}^{54} N(C_j)}, \quad (\sum_{j=1}^{54} P(C_j) = 1) \quad (3.3)$$

où $N(C_j)$ est le nombre d'observations appartenant à la classe C_j dans la base d'apprentissage et N la taille de la base.

Les nœuds x_i ($i = \overline{1,4}$) sont des nœuds enfants. Les probabilités calculées sont les probabilités conditionnelles de chaque variable x_i par rapport à son parent C_j . Elles sont calculées en utilisant la loi normale N , telle que $P(x_i/C_j) = N(x_i, \mu_{i,j}, \text{var}_{i,j})$ où $\mu_{i,j}$ et $\text{var}_{i,j}$ sont la moyenne et la variance du paramètre x_i dans la classe C_j .

Après avoir calculé les moyennes et les écarts-type de tous les paramètres de chaque classe à partir de la base de données, le réseau est maintenant quantifié. Ainsi, calculons des probabilités conditionnelles des nouvelles observations dans chaque classe.

3.5.3.3. Combinaison

Afin d'estimer la probabilité globale qu'une nouvelle observation appartienne à une classe C_j , il faut, dans un premier temps, combiner les différentes décisions multi-capteurs représentées par les probabilités

estimées dans l'étape précédente en utilisant la règle de décision de Bayes :

$$P(C_j/x_1, x_2, x_3, x_4) = \frac{P(C_j) \times P(x_1, x_2, x_3, x_4/C_j)}{P(x_1, x_2, x_3, x_4)} \quad (3.4)$$

où $P(C_j/x_1, x_2, x_3, x_4)$ est la probabilité *a posteriori* de la classe C_j conditionnellement aux paramètres x_1, x_2, x_3, x_4 , $P(C_j)$, la probabilité *a priori* de la variable classe C_j , $P(x_1, x_2, x_3, x_4/C_j)$, la vraisemblance des paramètres x_1, x_2, x_3, x_4 sachant la classe C_j , $P(x_1, x_2, x_3, x_4)$, la probabilité *a priori* des paramètres x_1, x_2, x_3, x_4 , appelé évidence ou encore constante de normalisation.

Le terme $P(x_1, x_2, x_3, x_4)$ dépend seulement des caractéristiques qui sont données. C'est un terme constant, facile à calculer et il est souvent négligé lors de la comparaison des différentes quantités $P(C_j/x_1, x_2, x_3, x_4)$, puisqu'il s'agit d'un nombre positif. Il est donné par l'expression (3.5).

$$\begin{aligned} P(x_1, x_2, x_3, x_4) &= \sum_{j=1}^{54} P(x_1, x_2, x_3, x_4/C_j) \times P(C_j) \\ &= \sum_{j=1}^{54} P(x_1/C_j) \times P(x_2/C_j) \times P(x_3/C_j) \times P(x_4/C_j) \times P(C_j) \end{aligned} \quad (3.5)$$

Cependant, le terme qui nous intéresse le plus est $P(x_1, x_2, x_3, x_4/C_j)$. Il est soumis à la loi de probabilités de plusieurs variables avec une forte indépendance des caractéristiques, appelée « *hypothèse d'indépendance naïve de Bayes* » et qui nous permet de factoriser le numérateur comme suit :

$$P(x_1, x_2, x_3, x_4/C_j) \times P(C_j) = P(x_1/C_j)P(x_2/C_j)P(x_3/C_j)P(x_4/C_j) \times P(C_j) \quad (3.6)$$

$$P(C_j/x_1, x_2, x_3, x_4) = \frac{P(C_j) \times P(x_1/C_j) \times P(x_2/C_j) \times P(x_3/C_j) \times P(x_4/C_j)}{P(x_1, x_2, x_3, x_4)} \quad (3.7)$$

Cette équation montre que le type de combinaison s'exprime sous la forme d'un produit, c'est, donc, une combinaison conjonctive. Elle est de plus associative. Sous l'hypothèse d'indépendance, elle devient commutative et associative.

Lorsque l'hypothèse d'indépendance est faite, l'approche bayésienne est qualifiée de naïve.

3.5.3.4. Décision

Le classificateur bayésien naïf est un modèle probabiliste basé, comme son nom l'indique, sur la règle de Bayes, qui dit qu'à tout élément $X = (x_1, x_2, x_3, x_4)$, on associe la classe C_j correspondant au maximum *a posteriori*, c'est-à-dire :

$$C_{MAP} = \operatorname{argmax}_{C_j} P(C_j/x_1, x_2, x_3, x_4) \\ \propto \operatorname{argmax}_{C_j} P(C_j) \times P(x_1/C_j) \times P(x_2/C_j) \times P(x_3/C_j) \times P(x_4/C_j) \quad (3.8)$$

Pendant, les classificateurs bayésiens naïfs peuvent utiliser d'autres stratégies d'apprentissage comme la règle majoritaire : à tout élément X , on associera la classe C_j telle que $P(C_j)$ est maximale ou la règle de la vraisemblance qui dit que, à tout élément X , on associera la classe C_j telle que $P(X/C_j)$ est maximale.

Comme exemple, prenons l'exemple de deux classes seulement $C_1 = \text{microclimat adéquat}$ et $C_2 = \text{microclimat inadéquat}$ (Tableau 3.6)

Tableau 3.6 Exemple de deux classes du microclimat

Etas du microclimat	La température ambiante (Ta)	La température du sol (Ts)	L'humidité ambiante (Ha)	Le rayonnement (Ra)
C_1	23.9529	20.1267	96.7245	141.2804
C_1	23.0466	20.0845	99.0183	101.4764
C_1	22.8598	21.0972	91.9916	170.2765
C_1	22.4143	18.4811	92.9828	150.6747
C_1	22.6468	20.5018	94.9652	138.1266
C_1	22.2217	19.3866	98.8990	106.4932
C_1	22.7504	19.3385	95.0141	135.8580
C_1	22.6598	20.2985	92.7700	123.4251
C_1	22.6842	21.4558	95.3396	120.3504
C_1	23.6342	18.7943	95.7425	181.3780
C_2	42.0980	25.3587	67.4863	30.9525
C_2	26.4645	67.6016	36.6588	68.7579
C_2	41.2524	29.5941	21.5543	98.6852
C_2	59.6497	35.7494	46.8765	76.9934
C_2	31.6037	54.9822	19.7170	82.9581
C_2	20.1376	-15.7894	75.8149	764.8678
C_2	-5.9549	0.5318	59.6640	676.2691
C_2	-5.4897	9.7384	73.4612	802.2993
C_2	-11.0136	18.8692	71.4490	597.3778
C_2	13.7574	10.2126	42.2195	892.1061

- Estimation des paramètres (Tableau 3.7)

Tableau 3.7 Calcul des moyennes et des variances

Classes C_j	$\mu_1(x_1)$	$var_1(x_1)$	$\mu_2(x_2)$	$var_2(x_2)$	$\mu_3(x_3)$	$var_3(x_3)$	$\mu_4(x_4)$	$var_4(x_4)$
C_1	22.8871	0.2840	19.9565	0.9159	95.3448	5.7662	136.9339	659.8511
C_2	21.2505	555.2334	23.6849	620.5510	51.4901	447.5953	409.1267	1.3257×10^5

$$P(x_1 = Ta/C_1) = N(x_1 = Ta, 22.8871, 0.2840)$$

$$P(x_3 = Ha/C_2) = N(x_3 = Ha, 51.4901, 447.5953)$$

Prenons une nouvelle observation à classer du tableau 3.8

Tableau 8 Une nouvelle observation à classer

Etat du microclimat	Ta (°C)	Ts (°C)	Ha (%)	Ra (Lux)
Inconnu	22,3	18	95,77	154,5

$$P(C_1) = P(C_2) = \frac{10}{20} = 0.5 \quad , \quad P(x_1 = 22.3/C_1) = 0.4080 \quad , \quad P(x_2 = 18/C_1) = 0.0516$$

$$P(x_3 = 95.77/C_1) = 0.1643, P(x_4 = 154.5/C_1) = 0.0123, P(x_1 = 22.3/C_2) = 0.0169$$

$$P(x_2 = 18/C_2) = 0.0156, P(x_3 = 95.77/C_2) = 0.0021, P(x_4 = 154.5/C_2) = 0.0009$$

- Combinaison

$$P(C_1/Ta, Ts, Ha, Ra) \propto P(C_1) \times P(Ta/C_1) \times P(Ts/C_1) \times P(Ha/C_1) \times P(Ra/C_1)$$

$$P(C_1/Ta, Ts, Ha, Ra) \propto 0.5 \times 0.4080 \times 0.0516 \times 0.1643 \times 0.0123 = 2.1257 \times 10^{-5}$$

$$P(C_2/Ta, Ts, Ha, Ra) \propto 0.5 \times 0.0169 \times 0.0156 \times 0.0021 \times 0.0009 = 2.4053 \times 10^{-10}$$

- Décision

Cependant, pour prendre une décision sur la nature du microclimat, il suffit de choisir la classe correspondant au maximum a posteriori donné par l'équation générale (3.8) et qui sera appliquée à notre cas comme suit :

$$C_{MAP} = \operatorname{argmax}_{C_j} P(C_j/x_1, x_2, x_3, x_4)$$

$$\propto \operatorname{argmax}_{C_j} P(C_j) \times P(x_1/C_j) \times P(x_2/C_j) \times P(x_3/C_j) \times P(x_4/C_j) \quad (3.9)$$

$$\text{Comme } P(C_1/Ta, Ts, Ha, Ra) > P(C_2/Ta, Ts, Ha, Ra) \quad (3.10)$$

Alors l'état du microclimat est, effectivement, adéquat pour l'olivier.

3.6. Evaluation du modèle développé

Pour évaluer le modèle établi en termes de taux de bonne classification et rapidité d'exécution, nous faisons un certain nombre de tests sur la base d'apprentissage et la base de test en appliquant deux algorithmes d'inférence différents pour pouvoir choisir le mieux adapté à la nature de nos données. Nous faisons varier aussi la taille de la base d'apprentissage. Nous calculons, à chaque fois, le taux et le temps de classification sur la base d'apprentissage et sur la base de test.

3.6.1. Tests sur la base d'apprentissage

Nous utilisons six échantillons de tailles respectives de 108, 540, 1080, 2700, 5400 et 16200, sur lesquels, nous avons appliqué l'algorithme d'apprentissage du maximum de vraisemblance « *learn_params* » et les deux algorithmes d'inférence l'arbre de jonction « *jtree_inf_engine* » et l'élimination des variables « *var_elim_inf_engine* ». Dans le but de conclure par rapport à la validité de notre méthode, nous calculons la matrice de confusion qui permet de donner le taux de bonne classification. Ce taux de classification ainsi que le temps d'apprentissage sont portés dans les tableaux 3.9 et 3.10.

- L'algorithme de l'arbre de jonction

Tableau 3.9 Résultats obtenus en appliquant l'algorithme de l'arbre de jonction sur la base d'apprentissage.

Taille de la base d'apprentissage	108	540	1080	2700	5400	16200
Taux de classification (%)	100	96.6667	90	89.5554	91.3333	91
Temps d'apprentissage (s)	3.847	7.459	15.041	35.431	70.327	212.019

- L'algorithme de l'élimination de variables

Tableau 3.10 Résultats obtenus en appliquant l'algorithme de l'élimination de variables sur la base d'apprentissage.

Taille de la base d'apprentissage	108	540	1080	2700	5400	16200
Taux de classification (%)	100	96.6667	90	89.5554	91.3333	91
Temps d'apprentissage (s)	8.521	12.476	22.783	53.889	75.633	498.447

Les résultats présentés, ci-dessus, montrent la capacité notable d'apprentissage de cette méthode. Même lorsque le nombre de données de la base d'apprentissage augmente, le taux de bonne classification reste satisfaisant avec un taux de bonne classification supérieur à 89%. Le temps d'apprentissage est plus grand lorsque la taille de la base est importante. Les deux algorithmes donnent exactement le même taux de classification pour les mêmes exemples. La différence réside dans le temps d'apprentissage qui est plus réduit avec l'algorithme de l'arbre de jonction.

En effet, l'algorithme le mieux adapté pour notre application est l'arbre de jonction.

3.6.2. Tests sur la base de test

Nous avons créé une base de test avec de nouveaux exemples. Pour étudier l'effet de la taille des données d'apprentissage sur la qualité de la méthode à classer de nouveaux exemples, nous avons choisi cinq échantillons d'apprentissage de tailles différentes de 162, 270, 540, 2700, et 5400 observations, en gardant la taille de la base de test constante à chaque fois. Dans chaque cas les tests ont été répétés 10 fois. Pour chacune des cinq tailles, nous avons calculé le taux de reconnaissance et le temps d'exécution en effectuant les moyennes des taux de classification et des temps d'exécution obtenus pour les 10 tests. Voici les résultats pour deux bases de test de tailles 54 et 162 observations.

Tableau 3.10 Résultats obtenus sur la base de test de 54 observations.

La taille de la base d'apprentissage	162	270	540	2700	5400
Le taux de classification (%)	66.6667	68.8888	77.7777	96.6666	100
Le temps d'apprentissage (s)	6.610817	10.258009	14.917541	50.688251	107.624209

Tableau 3.11 Résultats obtenus sur la base de test de 162 observations.

La taille de la base d'apprentissage	162	270	5400	16200
Le taux de classification (%)	56.2348	81,111	97.7778	100
Le temps d'apprentissage (s)	13.421408	15.566087	95.383454	286.266100

Nous constatons d'après les résultats des deux tableaux que le taux de bonne classification est en relation directe avec la taille de la base d'apprentissage. Plus la base est riche meilleur est le taux de classification. L'enrichissement de cette base reste indispensable pour avoir des taux d'erreurs faibles. Dans le contexte où nous travaillons, nous avons besoin de contrôler l'état du microclimat avec une période bien déterminée. Alors, le test se fait, réellement, sur une mesure à la fois.

Ainsi, avec une base d'apprentissage de 540 données, nous obtenons un taux de bonne classification de 100% pour les 54 états du réseau.

3.7. Conclusion

Dans ce chapitre, nous avons présenté les résultats obtenus de l'application des réseaux bayésiens à une fusion de données multi-capteurs pour le contrôle et la surveillance du microclimat d'un olivier cultivé sous serre. Cette méthode a donné de très bons résultats avec des taux de classification atteignant les 100 %, de plus, elle reste généralisable par rapport au nombre de paramètres utilisés et applicable pour n'importe quelle autre plante, à condition d'avoir les valeurs des paramètres les mieux adaptés pour sa croissance. Une fois l'état du microclimat est déterminé par le système développé, la prochaine étape est le contrôle des différents actionneurs afin d'apporter les changements nécessaires au niveau de chaque paramètre.

La structure naïve des réseaux bayésiens a permis la génération du réseau dans un temps réduit et obtenir des résultats de classification très satisfaisants, de plus cette approche se caractérise par sa souplesse et sa généralisation.

Conclusion générale

Le travail que nous avons présenté dans ce mémoire a été consacré à l'application de la technique d'apprentissage probabiliste des réseaux bayésiens sur une fusion de données multi-capteurs dans un contexte de classification pour le contrôle et la surveillance du microclimat d'une serre agricole afin de l'adapter à la culture d'olivier. Cette technique a fait preuve de robustesse et d'efficacité sur plusieurs types de données, complètes et incomplètes et prend en considération les incertitudes et les imprécisions des données délivrées par les différents capteurs. Les réseaux bayésiens sont l'une des méthodes les plus importantes de l'intelligence artificielle et restent préférables dans plusieurs cas car ils permettent : le rassemblement de plusieurs données de natures différentes dans un seul modèle, la représentation graphique qui reste compréhensible et facile à interpréter pas uniquement par des spécialistes contrairement aux systèmes experts par exemple, sans oublier qu'avec un réseau bayésien, on peut évaluer, prévoir, diagnostiquer ou optimiser des décisions.

Les étapes que nous avons suivies pour arriver à la fin de notre étude sont : la création de la base de données suivant les paramètres adéquats pour la meilleure croissance de l'olivier, modélisation du problème en utilisant les distributions de probabilités, estimation de ces probabilités et enfin la décision sur la nature du microclimat suivant les résultats donnés par le réseau.

Malgré la structure simple du réseau bayésien naïf que nous avons utilisé, les résultats obtenus sont très convaincants avec des taux de bonne classification atteignant les 100% et des temps de simulation réduits en utilisant l'algorithme d'inférence de l'arbre de jonction qui est le mieux adapté à la nature de nos données.

Ce travail nous a permis de s'ouvrir sur un certain nombre de perspectives :

Le premier point est d'étudier le positionnement et le placement de manière appropriée de ces capteurs et intégrer les autres paramètres climatiques comme le taux d'hygrométrie du sol, le taux de CO₂.

Le deuxième point est de commander les différents actionneurs pour contrôler chaque paramètre.

Le troisième point est la réalisation pratique dans une serre réelle.

Références bibliographiques

[1] Ayman ZUREIKI, « *Fusion de Données Multi-Capteurs pour la Construction Incrémentale du Modèle Tridimensionnel Texturé d'un Environnement Intérieur par un Robot Mobile* », Thèse de Doctorat. Université de Toulouse, 2008.

[2] David BELLOT, « *Fusion de données avec des réseaux Bayésiens pour la modélisation des systèmes dynamiques et son application en télémédecine* », 2005.

[3] M. Rombaut, « *Fusion : état de l'art et perspectives* », Laboratoire LM2S-UTT, Université de Troyes, 2001.

-
- [4] Isabelle Bloch, « *Fusion d'informations en traitement du signal et des images* », Hermes Science Publication, Janvier 2003.
- [5] P. Maybeck, « *The Kalmanfilter : An introduction to concepts* » New York, NY, USA. Springer-Verlag, 1990.
- [6] Ricco Rakotomalala, « *Pratique de la Régression Logistique - Régression Logistique Binaire et Polytomique* », version 2, Université Lumière Lyon 2, 2014.
- [7] Quinlan J.R., « *C4.5 : Programs for machine learning* » Morgan Kaufman Publishers, San Mateo, California, USA, 1993.
- [8] J.R. Quinlan, « *Induction of Decision Trees in Machine Learning* », Volume 1, 1986.
- [9] J. R. Quinlan, « *C4.5: Programs for Machine Learning* », Morgan Kaufmann Publishers, 1993.
- [10] David G. T. Denison, Bani K. Mallick and Adrian F.M. Smith, « *A Bayesian CART algorithm* », Department of mathematics, Imperial College, London, 1998.
- [11] G. Shafer, « *A mathematical Theory of Evidence* », Princeton University Press, 1976.
- [12] A. Dempster, « *Upper and lower probabilities induced by multi-valued mapping* », Annals of Mathematical Statistics AMS-38, 1967.
- [13] Isabelle Bloch, « *Fusion d'informations en image et vision* » Ecole Supérieure des télécommunications-CNRS UMR 5141 LTCI, Paris-France,
- [14] L. Zadeh, « *Fuzzy Sets, Information and Control* », 1965.
- [15] V. Barra, « *Fusion d'Images 3D du Cerveau : Etude de Modèles et Applications* ». Thèse de Doctorat. Université d'Auvergne, France, juillet 2000.
- [16] D. Dubois et H. Prade, « *Théorie des possibilités* », Masson, Novembre 1987.
- [17] M. Grabisch et P. Perny, « *Logique floue, principes, aide à la décision* », chapitre Agrégation multicritère, Hermes Science Publication, 2003.
- [18] B.V. Dasarathy, « *Sensor Fusion Potential Exploitation Innovative Architectures and Illustrative Applications* ». Proceeding of the IEEE, Janvier 1997.
- [19] D. Bellot, A. Boyer, and F. Charpillat, « *A new definition of qualified gain in a data fusion process: application to telemedicine* », Annapolis, Maryland, USA, 2002.
- [20] David Heckerman, « *A tutorial on learning with Bayesian network* », In Michael I. Jordan, editor, learning in Graphical Models, Kluwer Academic Publishers, Boston, 1998.
- [21] Paul Krause, « *Learning probabilistic networks* », 1998.
- [22] Patrick Naim, Pierre-Henri Wuillemin, Phillippe Leray, Olivier Pourret, Anna Becker, « *Réseaux Bayésiens* », Editions EYROLLES, 2007.
- [23] Judea Pearl and Tom Verma, « *A theory of inferred causation* », 1991.
- [24] Robert Cowell, A. Dawid, Steffen Lauritzen et David Spiegelhalter, « *Probabilistic Networks and Expert Systems. Statistics for Engineering and Information Science* », Springer-Verlag, 1999.
- [25] Radford Neal et Geoffrey Hinton, « *A view of the EM algorithm that justifies incremental, sparse and other variants* », In Michael Jordan, editor, Learning in Graphical Models, Kluwer Academic Publishers, Boston, 1998.

-
- [26] A. Dempster, N. Laird, et D. Rubin, « *Maximum likelihood from incomplete data via the EM algorithm* », Journal of the Royal Statistical Society, 1977.
- [27] Jie Cheng, David Bell et Weiru Liu, « *An algorithm for Bayesian network construction from data* », Proceedings of the sixth International Workshop on Artificial Intelligence and Statistics, 1997.
- [28] Jie Cheng, David Bell et Weiru Liu, « *Learning belief networks from data: An information theory based approach* », Proceedings of the sixth ACM International Conference on Information and Knowledge Management CIKM, 1997.
- [29] T. Verma et J. Pearl, « *Equivalence and synthesis of causal models* », in M. Henrion, R. Shachter, L. Kanal, et J. Lemmer, editors, Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence, San Francisco, Morgan Kaufmann, 1991.
- [30] P. Spirtes, C. Glymour et R. Scheines, « *Causality from probability* », in G. McKee ed. : *Evolving knowledge in natural and artificial intelligence*, London, 1990.
- [31] Judea Pearl, « *Bayesian networks: a model of self-activated memory for evidential reasoning* », In Technical Report 850021 (R-43), UCLA Computer Science Department Technical Report, and in Cognitive Science Society, UC Irvine, 1985.
- [32] Pearl, J., « *Fusion, propagation, and structuring in belief networks* », Journal of Artificial Intelligence, 1986.
- [33] S. Lauritzen et D. Spiegelhalter, « *Local computations with probabilities on graphical structures and their application to expert systems* », Journal of the royal statistical society, 1988.
- [34] F. Jensen, S. Lauritzen et K. Olsen, « *Bayesian updating in causal probabilistic networks by local computations* », Computational Statistics Quarterly, 1990.
- [35] Zhang, N. & Poole, D., « *A simple approach to bayesian network computations* », In Proceedings of the tenth Canadian Conference on Artificial Intelligence, 1994.
- [36] Dechter, R., « *Bucket elimination: a unifying framework for structure-driven inference* », Technical report, Department of Computer and Information Science, University of California, Irvine, USA, 1998.
- [37] Uffe Kjærulff, « *Approximation of Bayesian networks through edge removals* », Research Report IR-93-2007, Department of Computer Science, Aalborg University, Denmark, 1993.
- [38] Uffe Kjærulff, « *Reduction of computational complexity in Bayesian networks through removal of weak dependences* », In Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, San Francisco, California, 1994.
- [39] E. Horvitz, J. Suermondt, et G. Cooper, « *Bounded conditioning: Flexible inference for decisions under scarce resources* », In Proceedings of the fifth Conference on Uncertainty in Artificial Intelligence, North Holland, 1989.
- [40] Bruce D'Amboise, « *Incremental probabilistic inference* », In Proceedings of the ninth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1993.
- [41] Gilks, W., Richardson, S., & Spiegelhalter, D., « *Markov Chain Monte Carlo in Practice* », Interdisciplinary Statistics, Chapman & Hall, 1996.
- [42] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, et E. Teller, « *Equation of state calculations by fast computing machines* », Journal of Chemical Physics, 1953.

-
- [43] Radford Neal, « *Probabilistic inference using Markov chain Monte Carlo methods* », Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993.
- [44] Kevin Patrick Murphy, « *Dynamic Bayesian Networks: Representation, Inference and Learning* », université de Californie à Berkeley.
- [45] Roland Donat, Philippe Leray, Laurent Bouillaut et Patrice Aknin, « *Réseaux bayésiens dynamiques pour la représentation de modèles de durée en temps discret* », Journées francophone sur les réseaux bayésiens, 2008.
- [46] Kalman, R. E., Bucy R. S., « *New Results in Linear Filtering and Prediction Theory* », Transactions of the ASME - Journal of Basic Engineering Vol. 83, 1961.
- [47] Christophe Gonzales, « *Diagramme d'influence et CP-nets* », LIP6- Paris 6 Université, France, 2011.
- [48] E. Keogh et M. Pazzani, « *Learning augmented Bayesian classifiers: A comparison of distribution-based and classification based approaches* », In Proceedings of the Seventh International Workshop on Artificial Intelligence and Statistics, 1999.
- [49] N. Friedman, D. Geiger et M. Goldszmidt, « *Bayesian network classifiers* », Machine Learning, 1997.
- [50] Pat Langley, Wayne Iba, et Kevin Thompson, « *An analysis of Bayesian classifiers* », In Proceedings of the Tenth National Conference on Artificial Intelligence, AAAI Press, San Jose, CA, 1992.
- [51] J. Sacha, L. Goodenday et K. Cios, « *Bayesian learning for cardiac spect image interpretation* », Artificial Intelligence in Medicine, 2002.
- [52] Dan Geiger et David Heckerman, « *Knowledge representation and inference in similarity networks and Bayesian multinets* », Artificial Intelligence, 1996.
- [53] N. Friedman, D. Geiger, et M. Goldszmidt, « *Bayesian network classifiers* », Machine Learning, 1997.
- [54] S. Bezari, A. Bouhdjar et N. Ait Messaoudene, « *Etude du microclimat d'une serre tunnel équipée d'un dispositif de stockage thermique dans l'eau* », Revue des Energies Renouvelables CDER, Numéro Spécial ICRESD Tlemcen, 2007.
- [55] Gérard Houzard, « *Vers un classement des bioclimats des forêts caducifoliées françaises* », 1984.
- [56] Marie-Édith Tousignant et Michel Delorme, « *Connaitre le fonctionnement de la plante pour mieux gérer son environnement* », Article paru dans Québec Vert, septembre 2006.
- [57] V.B. Dasarathy, « *Sensor fusion potentiel exploitation-innovative architecture and illustrative applications* », Proc. Of IEEE, Volume 85, 1997.
- [58] M. R. Zemouri, « *Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamique, Application à la e-maintenance* », Thèse de doctorat, Université de Franche-Comté, 2003.
- [59] Didier MAQUIN, « *Surveillance des processus* », Institut National Polytechnique de Lorraine, Université HENRI POINCARÉ - NANCY 1, 2005.
- [60] Sylvie Charbonnier, « *Surveillance de systèmes continus à l'aide de méthodes n'utilisant pas de modèles formels : Application aux systèmes médicaux* », Université polytechniques de Grenoble, France, 2006.
- [61] <http://www.meteociel.fr>

