

Chapitre 3

Les réseaux Bayésiens

Présenté par :
Dr DJEBBAR Akila

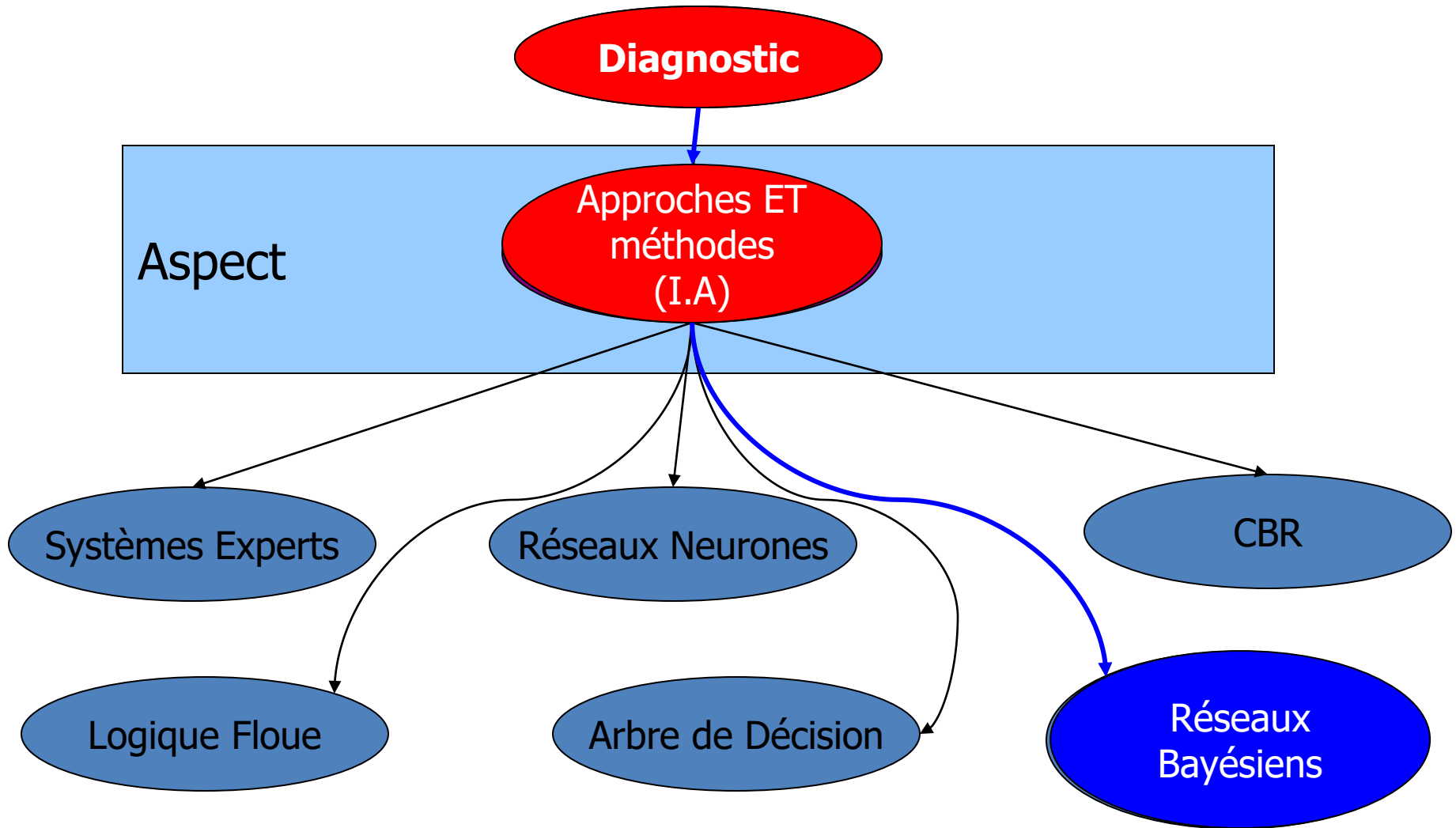
Plan du cours

- Introduction
- Réseaux Bayésiens : Définitions, propriétés,...
- Algorithmes d'inférences
- Exemples d'applications
- Conclusion

Introduction

- **L'un des buts principaux de l'intelligence artificielle (IA) est de concevoir des systèmes capable de reproduire le raisonnement humain.**
- **Les connaissances acquises ne sont pas toujours adéquates afin de permettre au système de faire une décision la plus appropriée possible.**
- **La notion de l'incertitude qui est intégré dans le raisonnement médical fait appel aux choix d'une approche probabiliste, ce qui nous amène tout naturellement au "Réseau bayésien" retrouvé par fois sous le nom de systèmes experts probabilistes.**

Approches en Diagnostic



Les Réseaux Bayésiens(1)

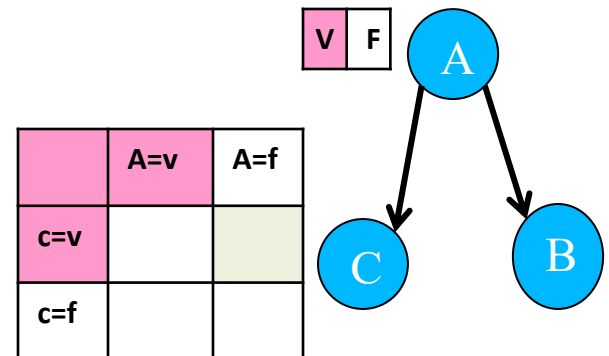
@ Definition:

Les réseaux bayésiens sont la combinaison des approches probabilistes et la théorie de graphes. Autrement dit, ce sont des modèles qui permettent de représenter des situations de raisonnement probabiliste à partir de connaissances incertaines.

@ Structure

1. Graphe **orienté** acyclique (DAG, **D**irected **A**cylic **G**raph).
2. Nœuds : ensemble de variables aléatoires (**discrete** ou **continue**).
3. Arcs **orientés**: des dépendances (causalités) probabilistes entre les variables et des distributions de probabilités conditionnelles pour chaque variable étant donné ses parents.
4. une table de probabilité pour chaque nœud.

Méthode symbolique et numérique



Les Réseaux Bayésiens(2)

- Méthodes statistiques
- Basés sur des probabilités conditionnelles
- La causalité joue un rôle important : des événements causent d'autres.
- Afficher graphiquement les variables d'un problème de décision et les influences entre ces variables.
- Basé sur le principe de théorème de Bayes.
-

Rappels sur les probabilités conditionnelles (1)

- ***Théorème de Bayes :***

Soient deux événements A et B . La probabilité a posteriori de l'événement A sachant la réalisation de l'événement B est :

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

De même la probabilité a posteriori de l'événement B sachant la réalisation de l'événement A est :

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

Rappels sur les probabilités conditionnelles (2)

On en déduit des deux formules précédentes que la probabilité conjointe des événements A et B est :

$$P(A \cap B) = P(B/A) \times P(A)$$

- ***Formule de Bayes :***

La probabilité a posteriori de l'événement A sachant la réalisation de l'événement B est :

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

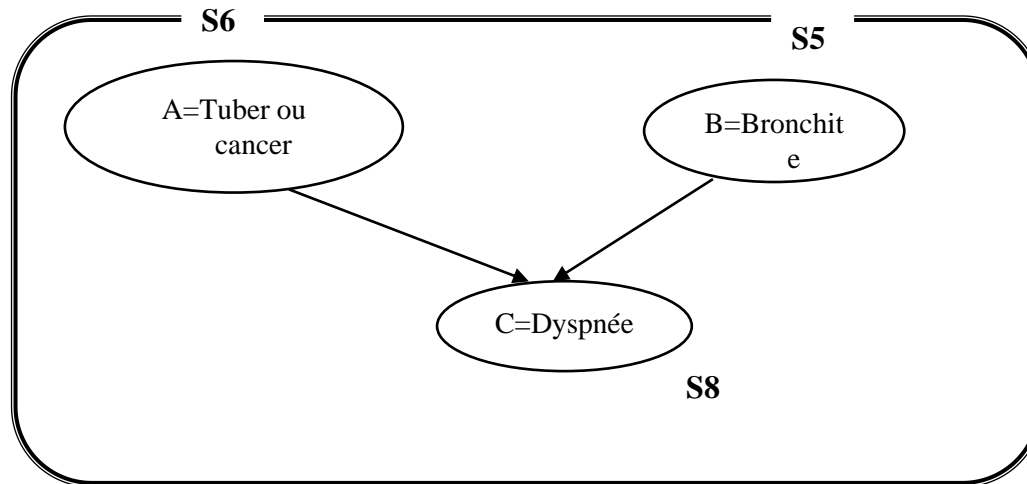
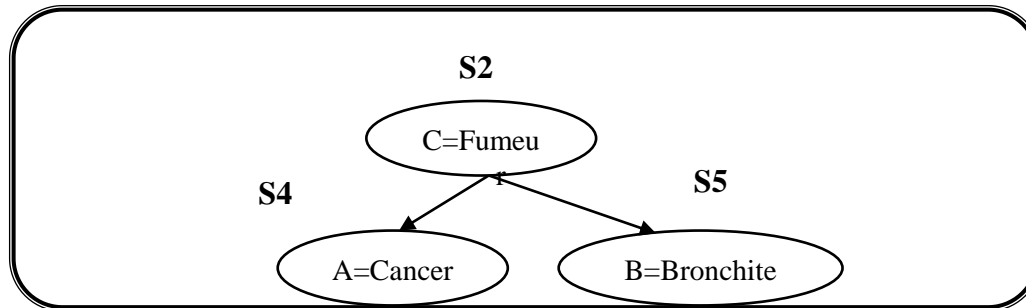
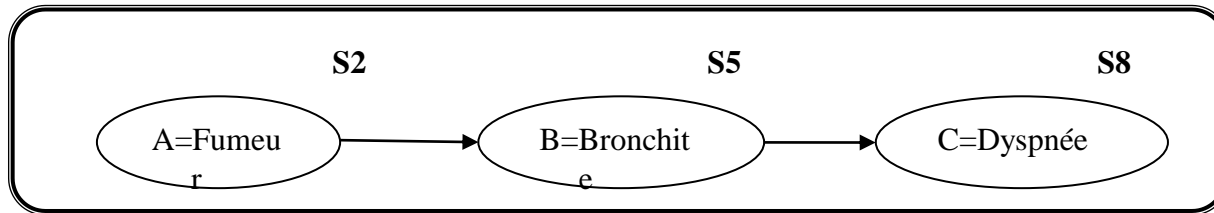
Indépendance conditionnelle (1)

Définition.

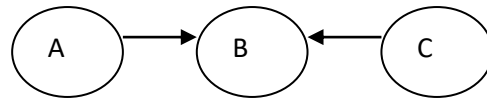
- La relation de définition de la **probabilité conditionnelle** de A lorsque B est réalisé, $P(A | B)$ = peut être écrite :
- $P(A | B) = P(A | B) P(B)$
- Si la probabilité de A n'est pas nulle, on peut écrire aussi :
- $P(A | B) = P(B | A) = P(B | A) P(A)$
- Deux événements A et B sont dits **indépendants en probabilité**, ou, simplement, **indépendants**, s'ils vérifient la relation :
- $P(A | B) = P(A) P(B)$
- Pour des événements de probabilité non nulle, les propriétés suivantes sont équivalentes :
 - A et B sont indépendants.
 - $P(A | B) = P(A) P(B)$.
 - $P(A | B) = P(A)$.
 - $P(B | A) = P(B)$.

L'information donnée par la réalisation de A sur la réalisation de B est nulle et l'information donnée par la réalisation de B sur la réalisation de A est nulle.

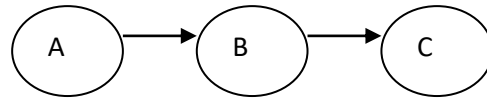
Indépendance conditionnelle (2)



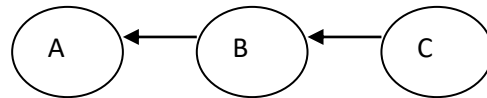
Circulation de l'information



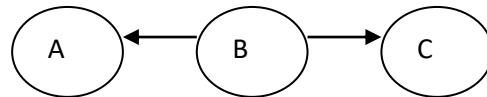
Connexion convergente
A et C causent B



Connexion en série
A cause B, B cause C



Cas symétrique



Connexion divergente
B cause A et C

Intuition de la classification bayésienne naïve

- **Objectif :**

- Associer à un document D (représenté par un vecteur de mots) la classe C_i qui lui correspond (parmi n catégories données : $\{C_1, C_2, \dots, C_n\}$).

Exemple : soit à classer une page dans l'une des catégories suivantes : {tourisme, culture, sport, science}

- **Mise en œuvre :**

- S'appuyer essentiellement sur la formule de Bayes pour estimer les probabilités suivantes :

$$P(C_1 / D), P(C_2 / D), \dots, P(C_n / D)$$

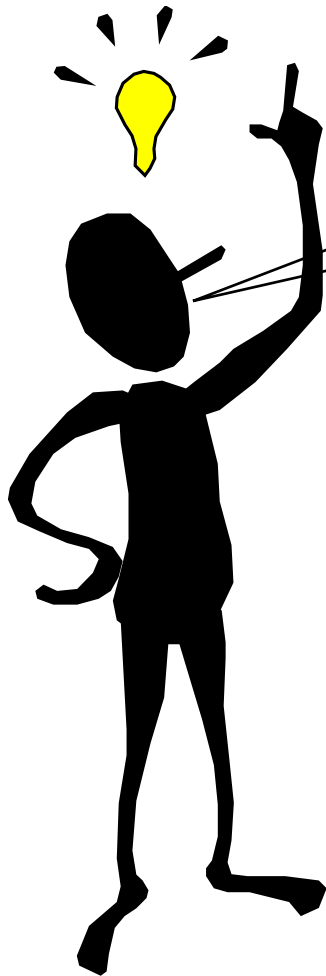
La classe choisie sera celle pour qui la probabilité est maximale

Question ?



Comment calculer pour
un document les
probabilités précédentes ?

Réponse :



Mais bien sûr : apprendre
à partir d'exemples que je
connais déjà !

Fondements de la classification bayésienne :

- a) Phase d'apprentissage*
- b) Phase de test*

Phase d'apprentissage

- Hypothèses & Notations :

- L'ensemble C des classes de classification :

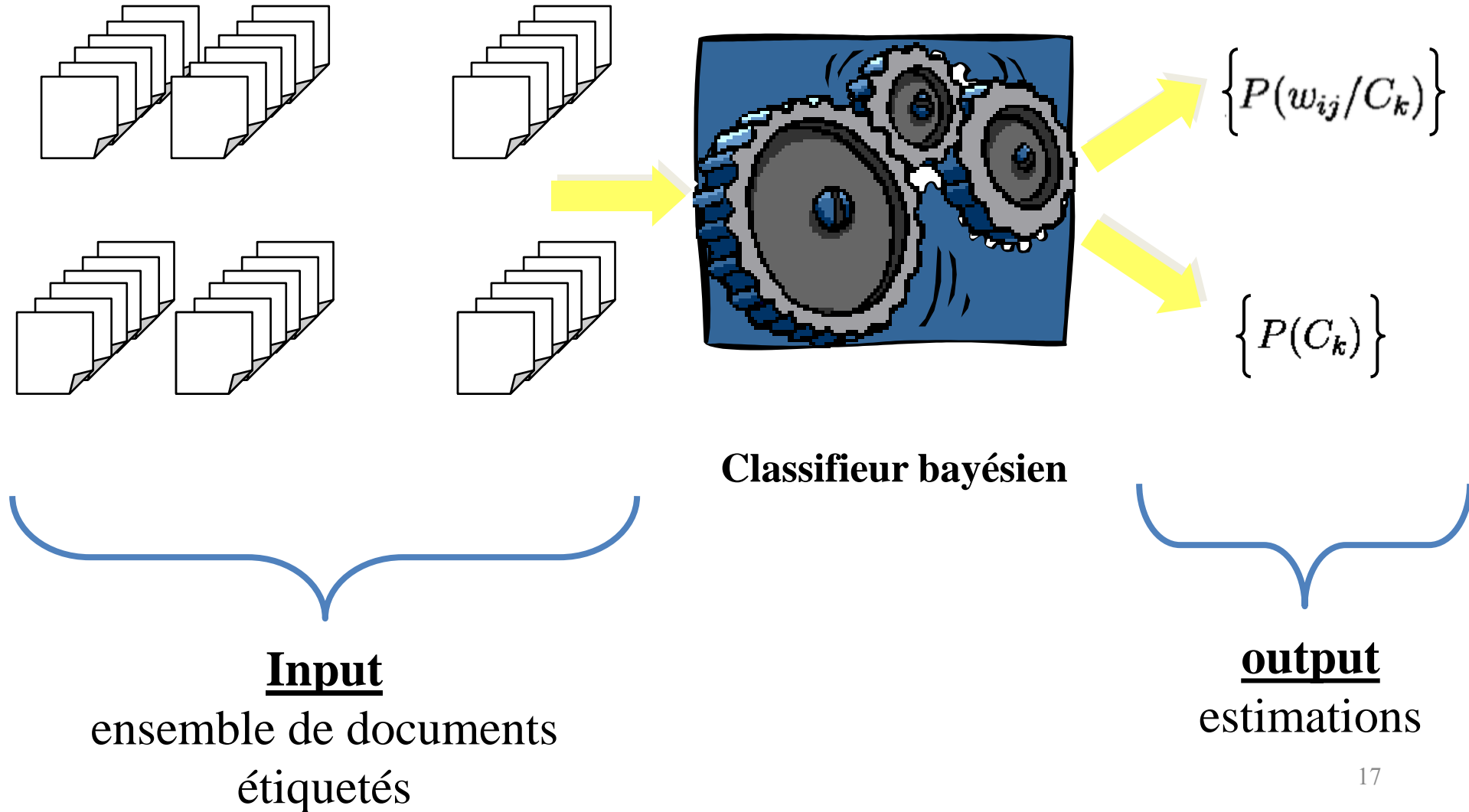
$$C = \{C_k, k = 1..|C|\}$$

- Un modèle paramétrique théorique $M(\theta)$ noté M permettant de générer, à partir d'un vocabulaire V , des documents conformes aux classes précédentes.

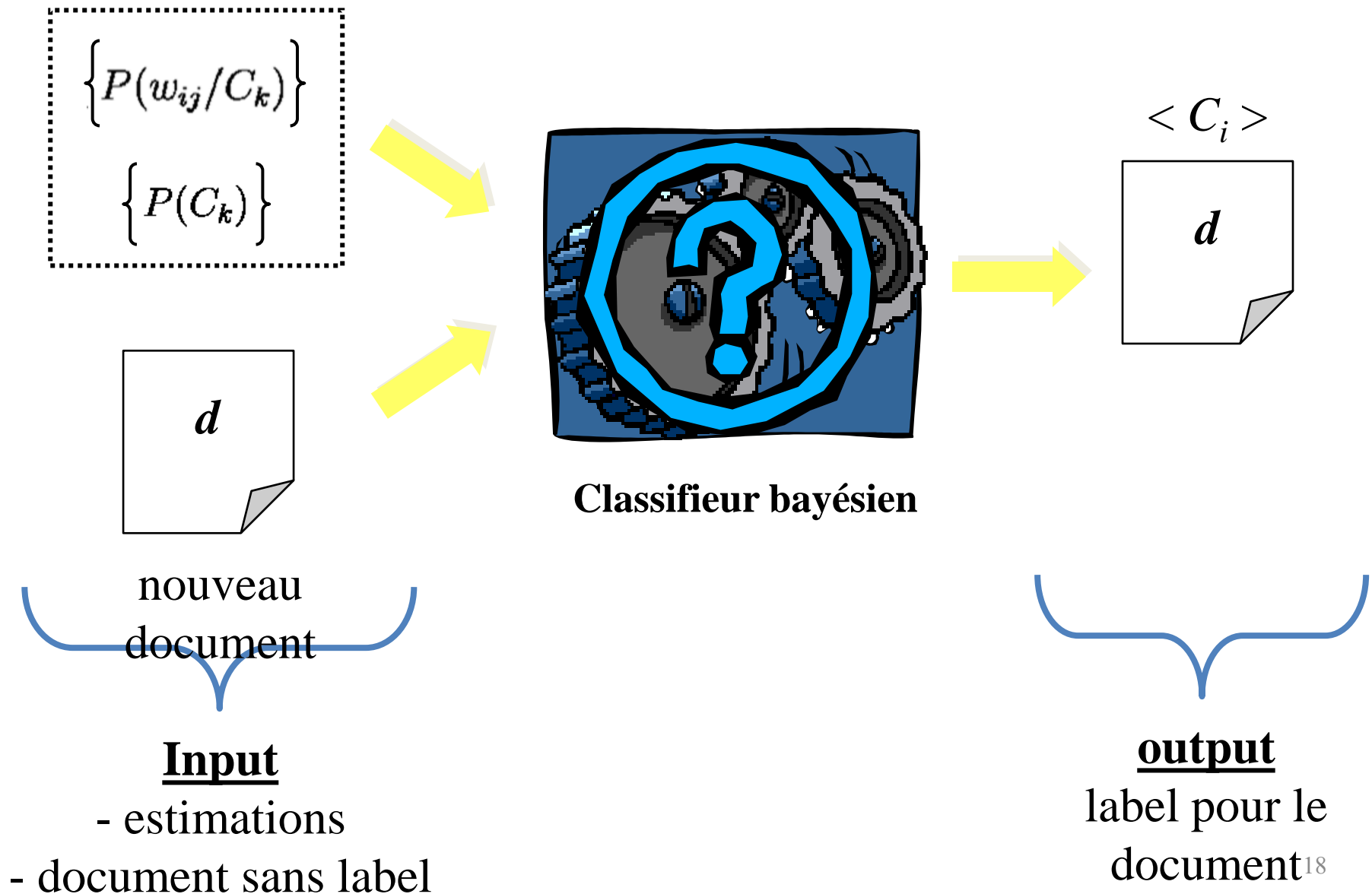
- Le modèle M associe à chaque document d_i un certain format v_i appelé ***vecteur caractéristique***.

$$\mathcal{M} : d_i \longrightarrow v_i = \langle w_{i1}, w_{i2}, \dots, w_{i|v_i|} \rangle$$

Phase d'apprentissage



Phase de test



Phase de test

- Soit d_i un document dont v_i est le vecteur caractéristique associé par le modèle M :

$$P(C_k/v_i) = P(C_k) \times \frac{P(v_i/C_k)}{P(v_i)}$$

- **Hypothèse naïve du classifieur :**

Indépendance d'occurrence des mots d'un document :

- Après simplification $P(v_i/C_k) = \prod_{j=1}^{|v_i|} P(w_{ij}/C_k)$

$$P(C_k/v_i) \propto P(C_k) \times \prod_{j=1}^{|v_i|} P(w_{ij}/C_k)$$

Phase de test

- **La classification :**

Choisir la classe la plus probable pour générer un tel document

$$C_p = \arg \max_{C_k} P(C_k/v), k = 1..|C|$$

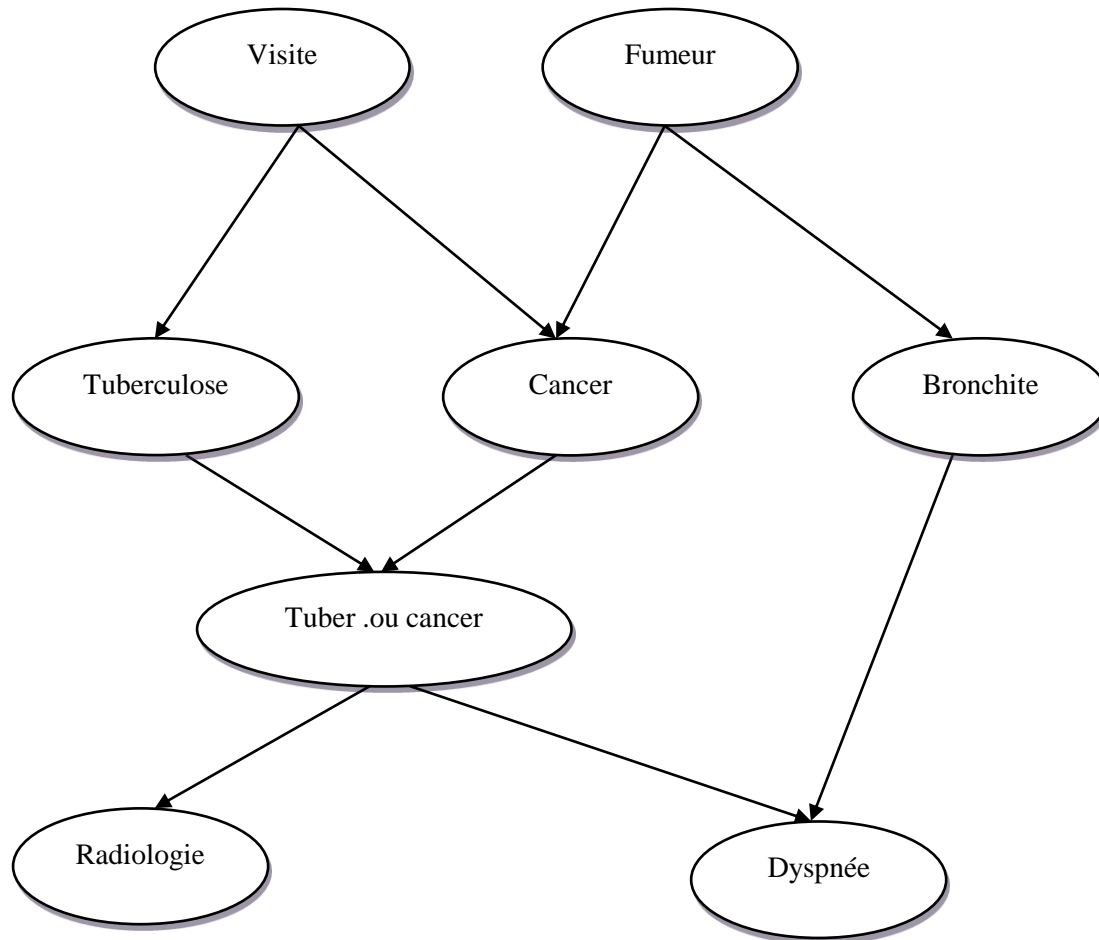
L'incertain dans le domaine médical

- Le diagnostic = identification d'une maladie par ses symptômes
- La décision médicale est toujours associée à **un degré d'incertitude** qui est inhérent au domaine de la biologie et de l'humain, ceci est du à la difficulté et l'impossibilité de :
 - De recueillir certaines données physiopathologiques
 - Et de faire des mesures et examens sans déroger à la déontologie.
 - La rareté de certaines ressources
 - ...

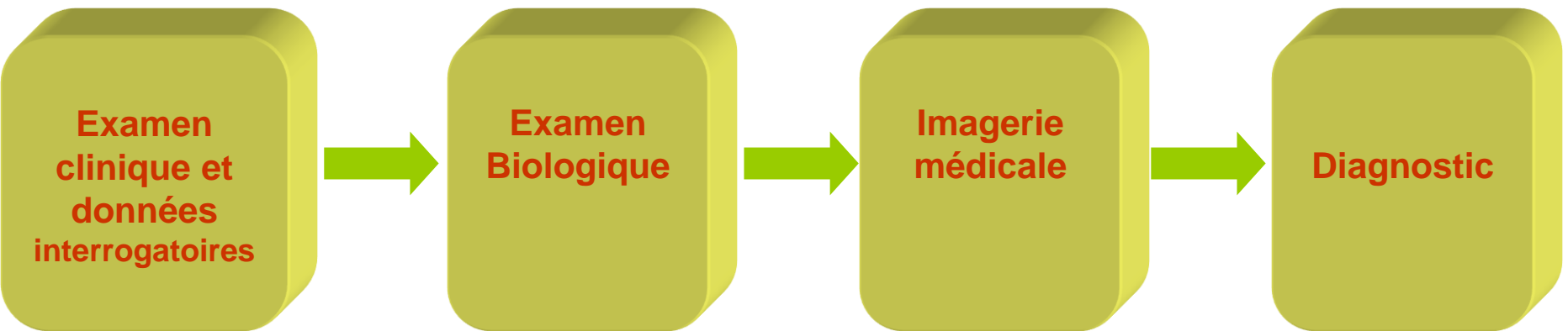
Il existe une variabilité d'inter-sujet

Le processus décisionnel de diagnostic est un Processus sous incertitude

Exemple d'un réseau bayésien dans le domaine médical



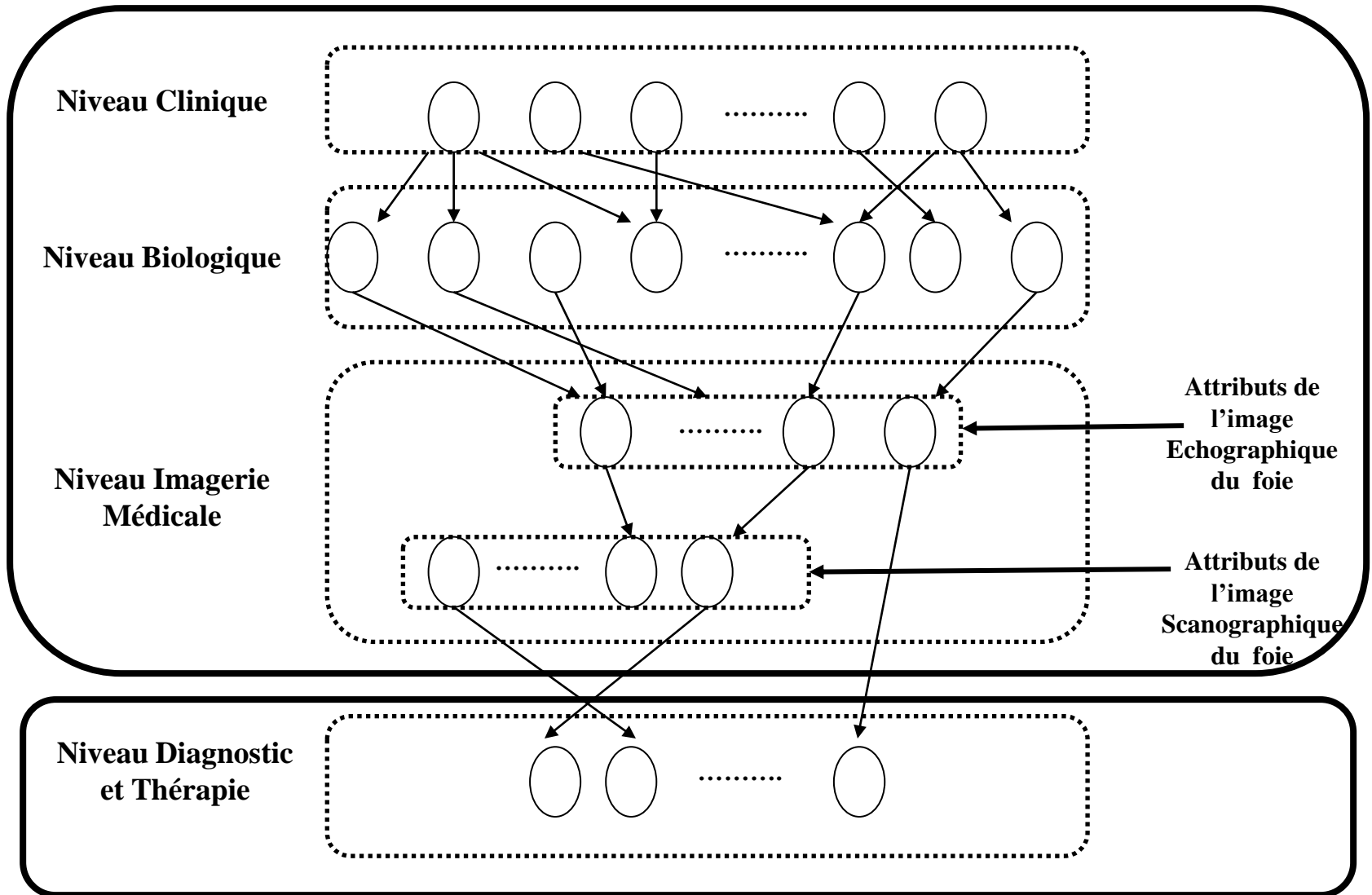
Description générale d'un problème médical



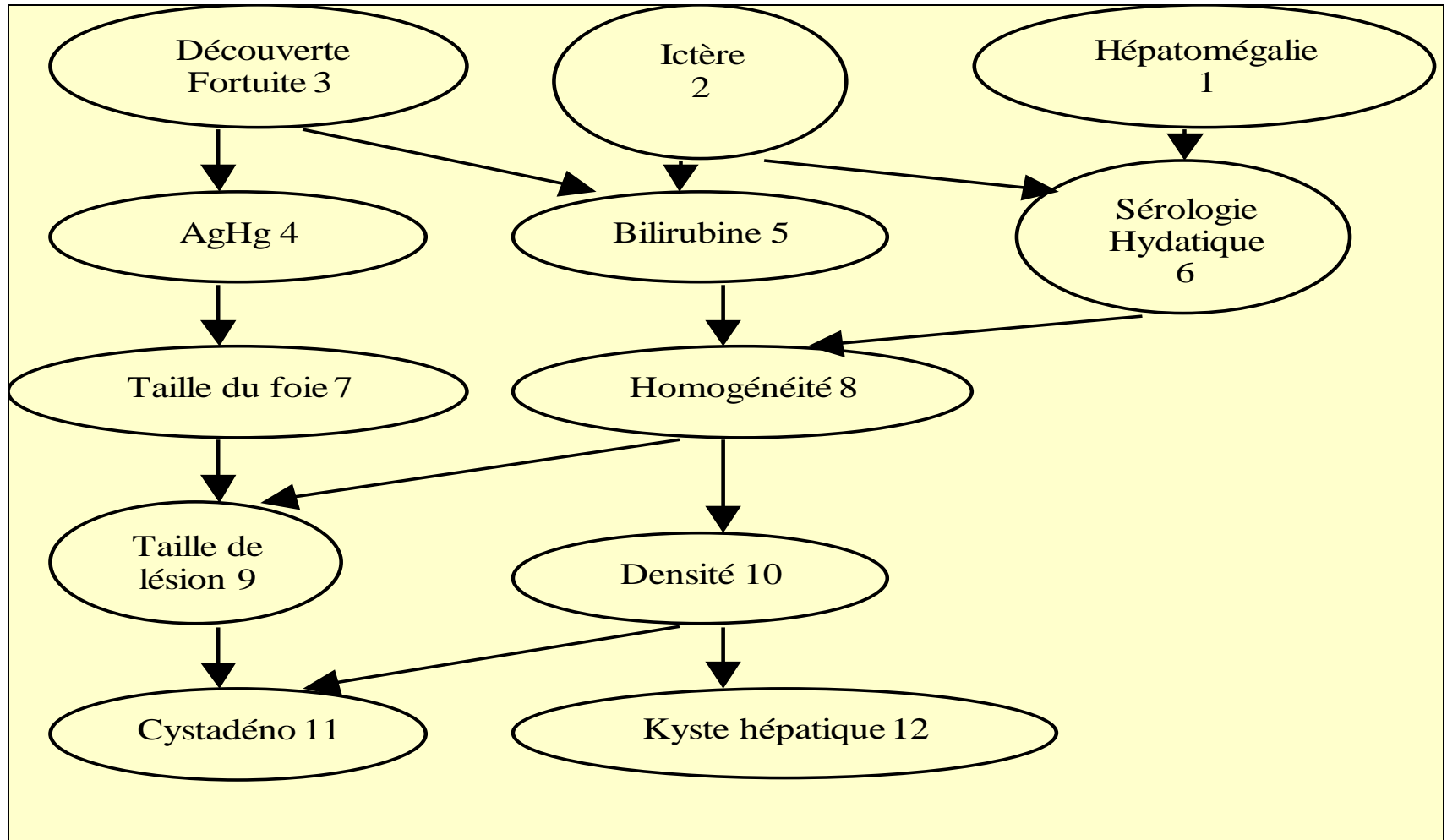
Algorithmes d'inférences

- JLO (Junction Tree)
- PEARL (message Passing)
- EM(Expectation Maximisation)
-

Modélisation d'un problème médical par un réseau Bayésien



Exemple de réseau Bayésien avec deux maladies du foie



Exemple de réseau Bayésien avec deux maladies du foie

Description d'un cas

Partie Clinique:

Douleur=1

Ictère=0

Tabac=1

Partie Biologique:

.....

Imagerie médicale

Attributs de l'image échographique:

Taille=..

Homogénéité=..

Attributs de l'image scanographique:

Taille=..

Densité=..

Angle Aigu=..

.....

Solution (diagnostic-Thérapie):

Cirrhose-Fonction Biopsie du foie



Algorithme JLO

Relier les parents entre eux et les

JLO est obtenu à partir du graphe triangulé en connectant les cliques de telle façon que toutes les cliques sur le chemin entre deux cliques X et Y contiennent $X \cap Y$.

Ajout sélectif des arcs au graphe moral pour formé un graphe triangulé.

Regularisation

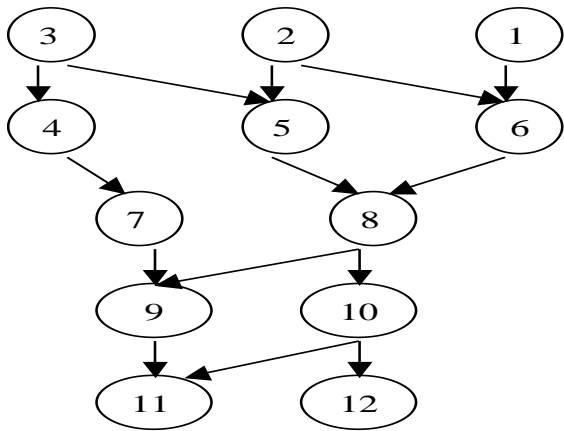
Regularisation

Arbre de jonction

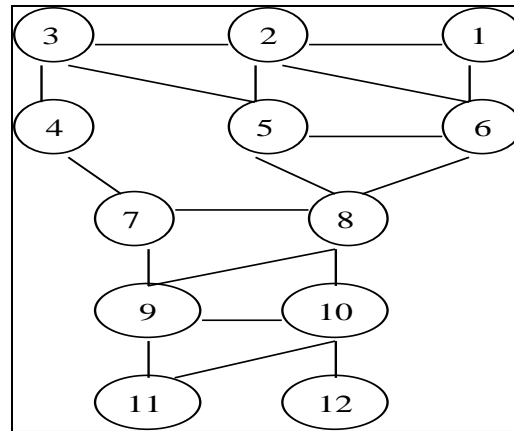
2. Calculs des fonctions potentiels associées à l'arbre de Jonction

3. Etape de propagation

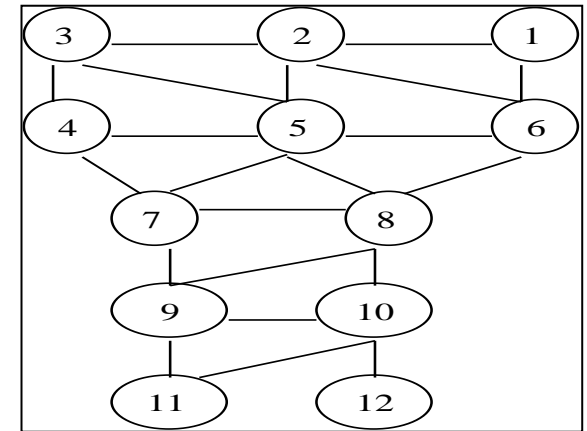
■ la recherche de probabilités se réalise dans l'arbre de Jonction



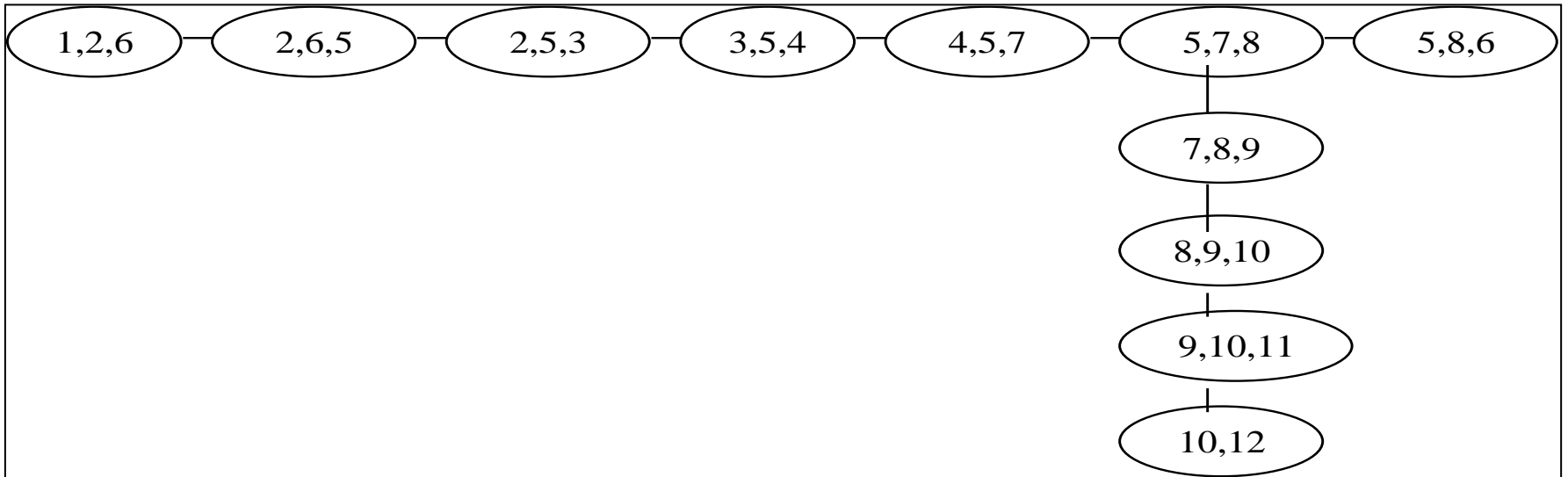
(a) Réseau Bayésien



(b) Moralisation



(c) Triangulation



(d) Arbre de Jonction

Exemple de construction d'un arbre de Jonction à partir d'un réseau bayésien de deux maladies du foie

Algorithme JLO : Discussion

■ **l'algorithme JLO permet :**

- **construire un graphe (sans boucle) plus efficace**
- **plus rapide**
- **réduire la complexité en temps et en espace mémoire**

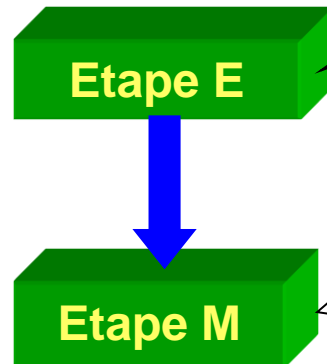
1. Inférence : Algorithme EM

Expectation Maximisation par Dempster 1977

Le principe de l'algorithme EM :

➔ Traiter des bases d'exemples incomplètes

➔ Éviter d'ajouter une nouvelle modalité



Estimation des variables manquantes

Utilisation des variables connues estimées pour estimer de nouveaux paramètres jusqu'à convergence.

Les étapes de l'algorithme EM

Algorithme EM : Exemple

Hépatomégalie	Ictère	Douleur
Grand	Oui	Oui
Grand	Oui	Non
-	Non	Oui
Petit	-	Non
-	Oui	Non
Petit	Non	Oui
Moyen	Oui	Non
-	Non	Oui

Iteration 1

H	I	D	P(H=G)	P(H=P)	P(H=M)
G	O	O	1	0	0
G	O	N	1	0	0
-	N	O	0.333	0.333	0.333
P	-	N	0	1	0
-	O	N	0.333	0.333	0.333
P	N	O	0	1	0
M	O	N	0	0	1
-	N	O	0.333	0.333	0.333
Total			3	3	2

Iteration 2

H	I	D	P(H=G)	P(H=P)	P(H=M)
G	O	O	1	0	0
G	O	N	1	0	0
-	N	O	0.375	0.375	0.250
P	-	N	0	1	0
-	O	N	0.375	0.375	0.250
P	N	O	0	1	0
M	O	N	0	0	1
-	N	O	0.375	0.375	0.250
Total			3.125	3.125	1.75

Algorithm EM : Exemple

Iteration 3

H	I	D	P(H=G)	P(H=P)	P(H=M)
G	O	O	1	0	0
G	O	N	1	0	0
-	N	O	0.390	0.390	0.219
P	-	N	0	1	0
-	O	N	0.390	0.390	0.219
P	N	O	0	1	0
M	O	N	0	0	1
-	N	O	0.390	0.390	0.219
Total			3.17	3.17	1.657

Iteration 4

H	I	D	P(H=G)	P(H=P)	P(H=M)
G	O	O	1	0	0
G	O	N	1	0	0
-	N	O	0.396	0.396	0.207
P	-	N	0	1	0
-	O	N	0.396	0.396	0.207
P	N	O	0	1	0
M	O	N	0	0	1
-	N	O	0.396	0.396	0.207
Total			3.188	3.188	1.621

Iteration 5

H	I	D	P(H=G)	P(H=P)	P(H=M)
G	O	O	1	0	0
G	O	N	1	0	0
-	N	O	0.398	0.398	0.202
P	-	N	0	1	0
-	O	N	0.398	0.398	0.202
P	N	O	0	1	0
M	O	N	0	0	1
-	N	O	0.398	0.398	0.202
Total			3.194	3.194	1.606

Iteration 6

H	I	D	P(H=G)	P(H=P)	P(H=M)
G	O	O	1	0	0
G	O	N	1	0	0
-	N	O	0.399	0.399	0.200
P	-	N	0	1	0
-	O	N	0.399	0.399	0.200
P	N	O	0	1	0
M	O	N	0	0	1
-	N	O	0.399	0.399	0.200
Total			3.197	3.197	1.600

Algorithme EM : Discussion

- **l'arbre de JLO est incapable de donner une classification en cas de présence des valeurs manquantes**
- **Evite d'enrichir la base de cas par des informations non mesurée (non pertinentes)**



Afin d'améliorer les performances de notre système

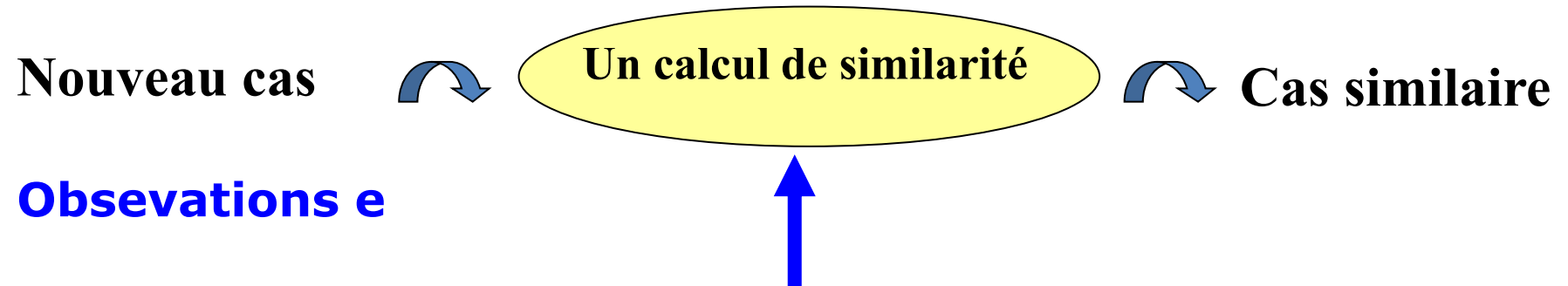
2. Recherche

- Elle consiste à déterminer le diagnostic associé à ce nouveau cas en se basant sur les probabilités (obtenus par l'arbre de Jonction) comme des mesures de similarité.
- A partir des observations e_i , on cherche :

$$\arg \text{Max} (P(C=c_i | e))$$

C : est une variable aléatoire qui peut prendre plusieurs valeurs correspondant à des ensembles de cliques c_i .

Recherche



La clique la plus probable

$$\arg \text{Max} (P(C=c_i | e))$$

C : est un nœud de l'arbre de Jonction (Clique)

e : nouveau cas

4. Mémorisation

Pour chaque clique contenant la variable v recevant l'évidence, la fonction potentiel de la clique $\emptyset C$ de la clique est multipliée par l'évidence.

Nouveau cas

Ictère (1)

1,7,
8

$$Fct1 = Fct1 * P(1)$$

Biliribune (2)

Sérologie hédatique (3)

2,3,4

$$Fct2 = Fct2 * P(2) * P(3)$$

Douleur (4)

2,5,
6

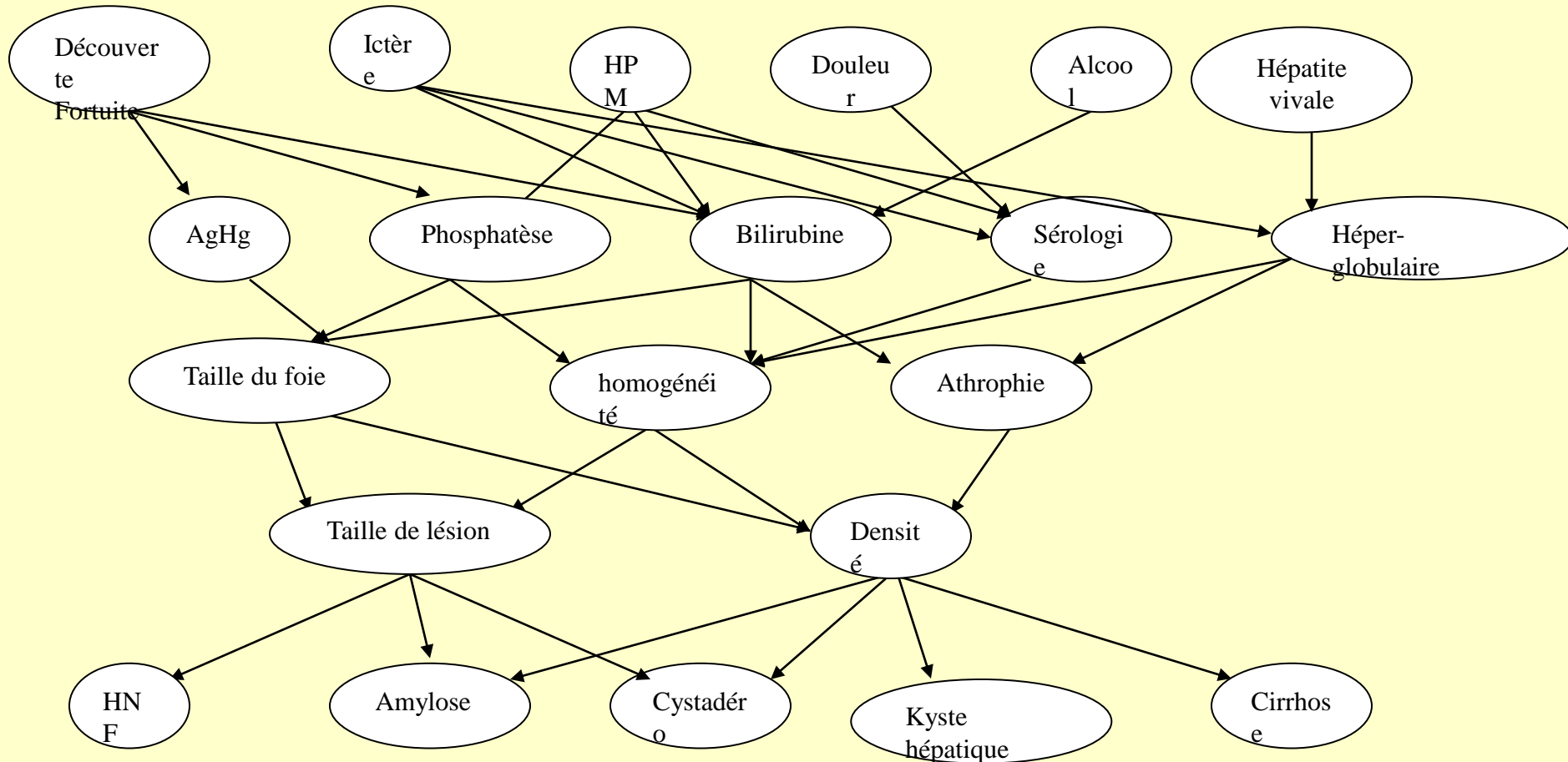
$$Fct3 = Fct3 * P(2)$$

.....

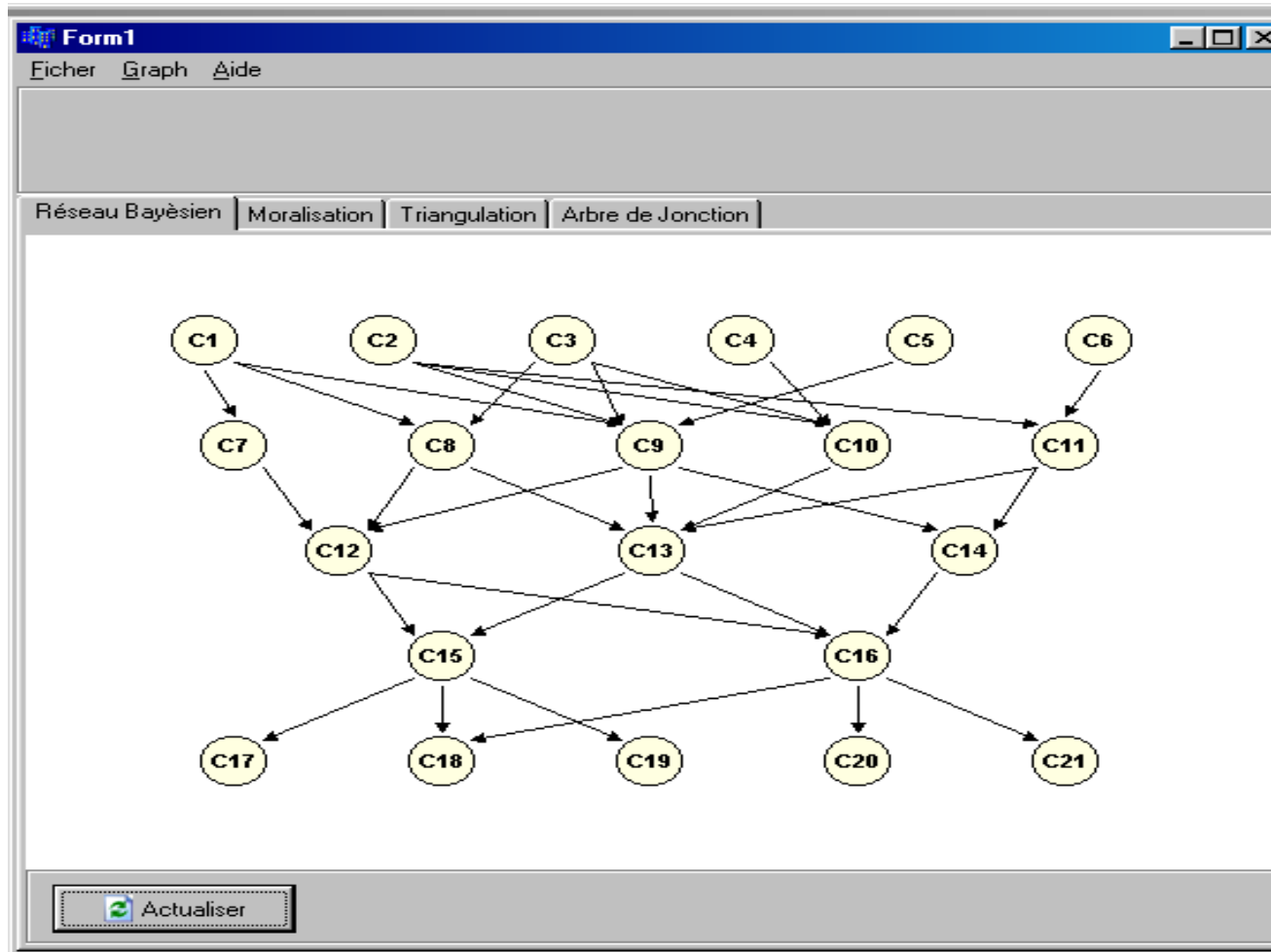
Résultats Expérimentaux

- **Test sur une base de 5 cas**
- **Sous Builder C++ version 6.0**
- **Réseau de 21 nœuds :**
 - **16 nœuds = représentant les signes de trois niveaux (clinique, biologique, imagerie médicale)**
 - **5 nœuds = représentant le niveau diagnostic et thérapie.**

Réseau Bayésien modélisant une de base de cas de 5 cas

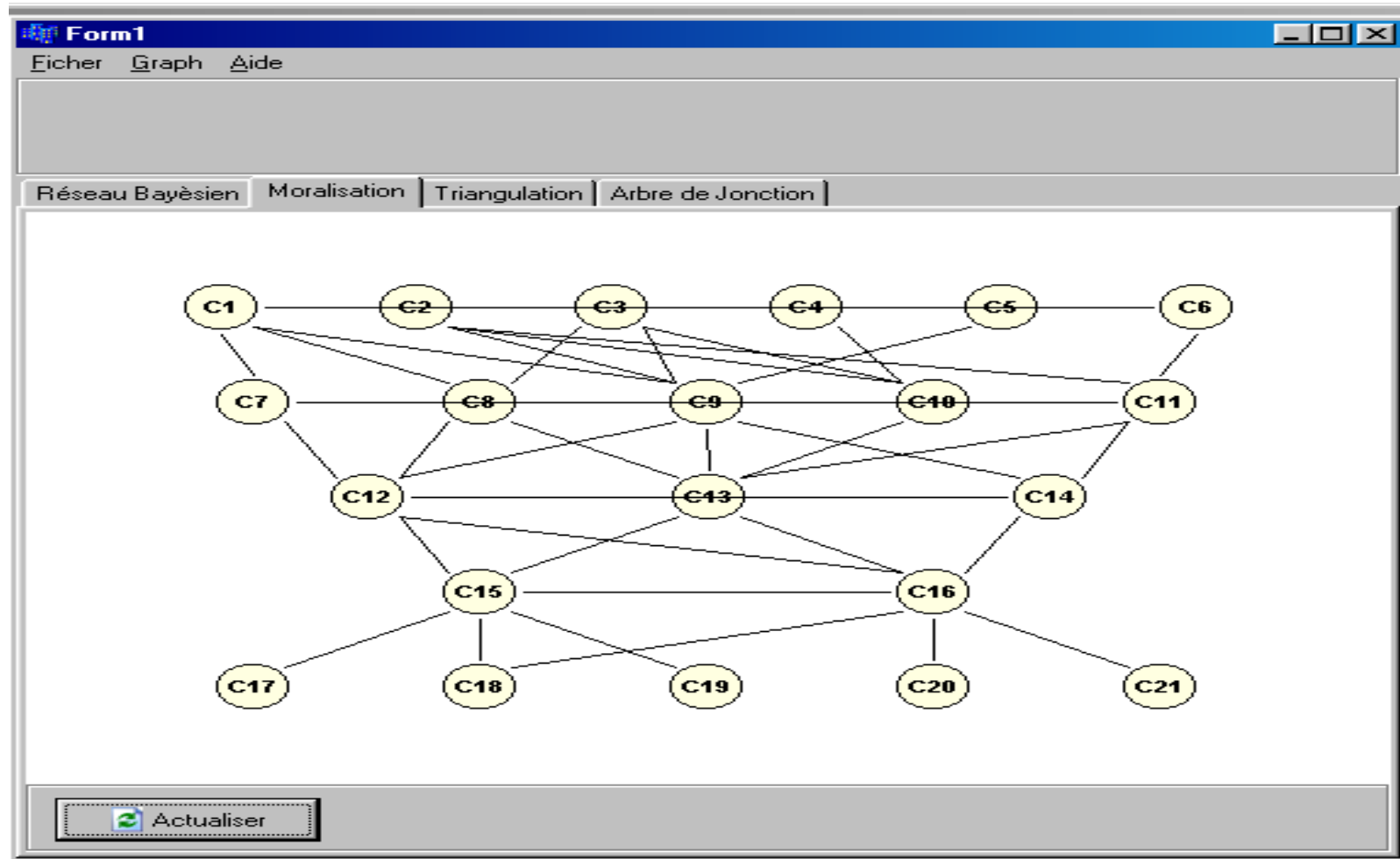


Construction de l'arbre de jonction JLO



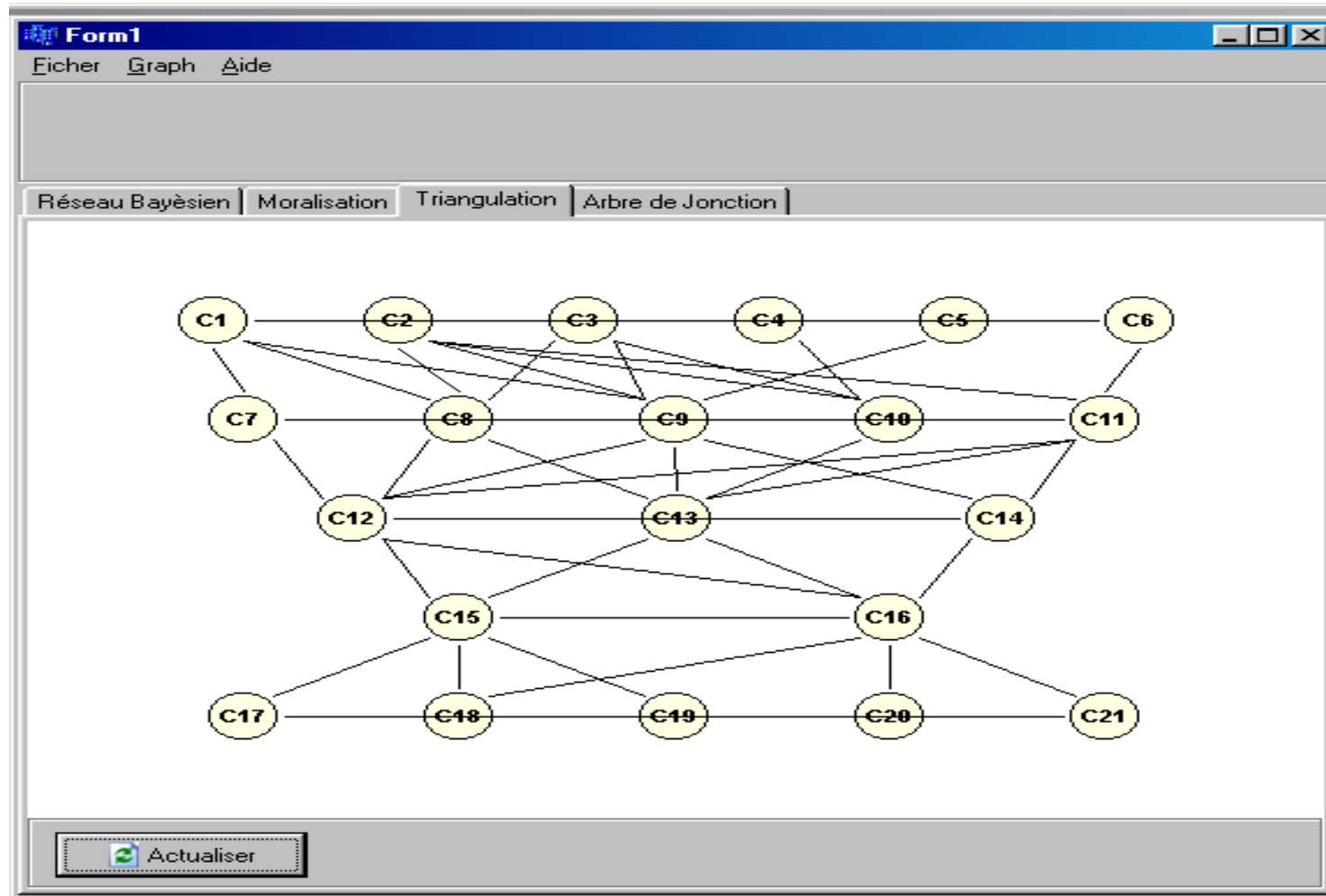
Construction de l'arbre de jonction JLO :

1. Etape de moralisation



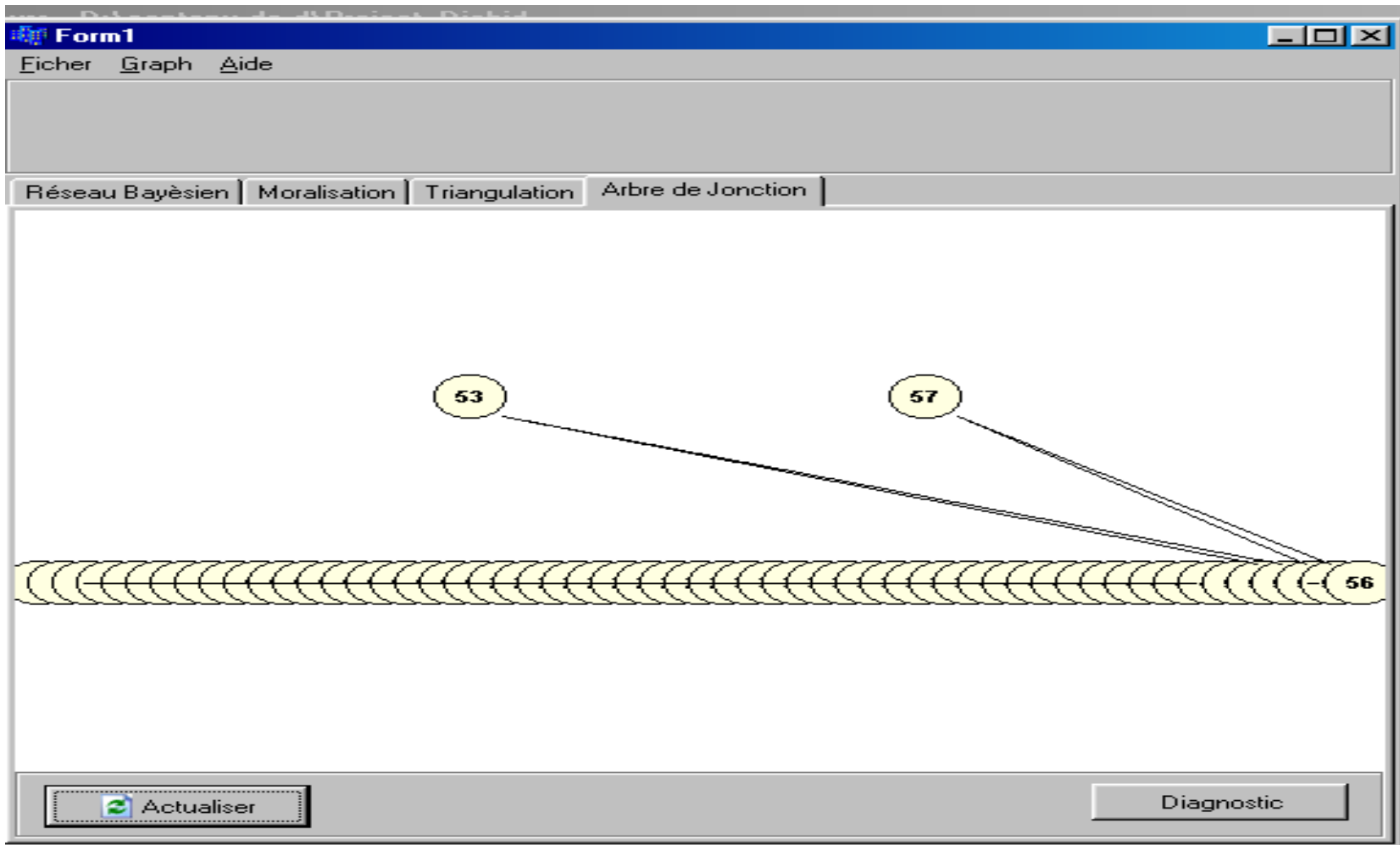
Construction de l'arbre de jonction JLO :

2. Etape de triangulation




Construction de l'arbre de jonction JLO :

3. Arbre de Jonction



Classification du nouveau cas 1/2

Form2



Signes cliniques

- Découverte fortuite
- Ictère
- HPM
- Douleur
- Alcool
- Hépatite virale

Examen biologique

- AgHg
- PH
- Bilirubine
- Sérologie
- Hyperglobuline

Imagerie médicale

- Taille du foie
- Homogénéité
- Atrophie
- Taille de lésion
- densité

Mise à jour Traitement

Classification du nouveau cas 2/2

ViewDetailForm [-] [□] [X]

Fonction potentielle :

Diagnostic :

Taux de similarité :

Thérapie :

1. Si (la lésion est unique ou multiples) alors Traitement=Chirurgie
2. Si (la lésion est de petite taille) alors Traitement=Transplantation
3. Si (la lésion est inopérable) alors Traitement=Chémioembolisation

Détection d'un nouveau cas

The screenshot shows a software window titled "Form2" with a header image of a hand holding a globe. The main area contains three columns of checkboxes:

- Signes cliniques**
 - Découverte fortuite
 - Ictère
 - HPM
 - Douleur
 - Alcool
 - Hépatite virale
- Examen biologique**
 - AgHg
 - PH
 - Biliribune
 - Sérologie
 - Hyperglobuline
- Imagerie médicale**
 - Taille du foie
 - Homogénéité
 - Atrophie
 - Taille de lésion
 - densité

An "Information" dialog box is overlaid in the foreground, containing the text "Détection d'un nouveau Cas" and an "OK" button.

Possibilité d'ajouter un nouveau nœud au réseau initial

DirectedGraphForm

Directed Acyclic Graph.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	
C1							X	X	X													
C2									X	X	X											
C3								X	X	X												
C4									X													
C5								X														
C6										X												
C7											X											
C8											X	X										
C9											X	X	X									
C10												X										
C11												X	X									
C12														X	X							
C13														X	X							
C14															X							
C15																X	X	X				
C16																	X		X	X		
C17																						

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21
W	1	1	1	1	1	1	2	2	2	2	2	3	3	3	4	4	5	5	5	5	5
P	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1

Ajouter

Etiquette:

Poids:

Probabilité:

Valider

● **Algorithme de propagation:**

1) *Calcul de la probabilité actuelle*

La
probabilité
de X
sachant
ses parents

Message
venant de
 Z à son fils
 X

$$P(X=x) = \pi(X=x) = \sum_z P(X=x | Z=z) \pi_x(Z=z)$$

2) *Construction de liste des fils*

3) *Diffuser un message π contenant la probabilité calculée à tous les fils*

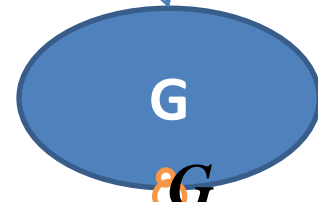
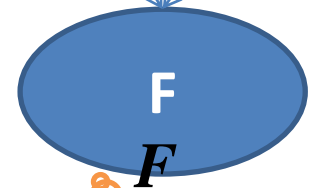
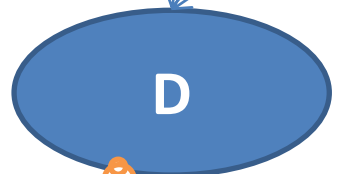
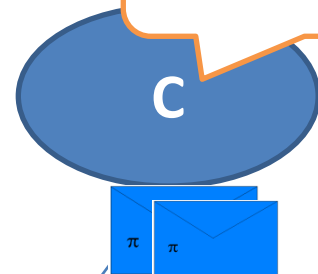
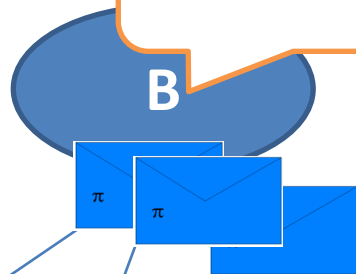
4) *Les fils recalculent ses nouvelles probabilités à la lumière de message reçu*

Propagation du réseau

Calcul de π (A)

Calcul de π (B)

Calcul de π (C)



D

E

F

G

*recalcul
e sa
probabil
ité*

*recalcul
e sa
probabil
ité*

*recalcul
e sa
probabil
ité*

*recalcul
e sa
probabili
té*

Algorithme de Pearl

- Le modèle graphique est simple classé comme un polyarbres avec plusieurs racine.
- Raisonnement avec données complètes d'où le principe de l'inférence exacte de Pearl.
- Algorithme de Pearl est un algorithme exact qui permet de parcourir tous l'espace de recherche de notre graphe.
- La simplicité du calcul de la loi marginale d'une variable ou de sa loi conditionnelle relativement à un ensemble d'observations.

Exemple d'application d'un réseau bayésien Naïf (1/5)

Apprendre à reconnaître un visage dans une image

Utilisation d'un réseau bayésien naïf pour reconnaître un visage dans une image, à partir de la couleur de sa peau

1. Modélisation : on va utiliser un réseau bayésien naïf qui va définir le lien entre les points du visage (et ceux qui n'appartiennent pas au visage) et la distribution de probabilités sur les couleurs de ces points;
2. Apprentissage : à partir d'un ensemble d'images, on estime les tables de probabilités;
3. Inférence : on se sert du réseau bayésien naïf pour déterminer pour chaque point sa probabilité d'appartenir au visage ou pas (sachant les informations sur la couleur du point)



e-Motion Group ©2007
Olivier.aycard@imag.fr

 UNIVERSITÉ JOSEPH FOURIER
LEZARD - DÉPARTEMENT - MONTAIGNE - JARDIN - VILLE

29

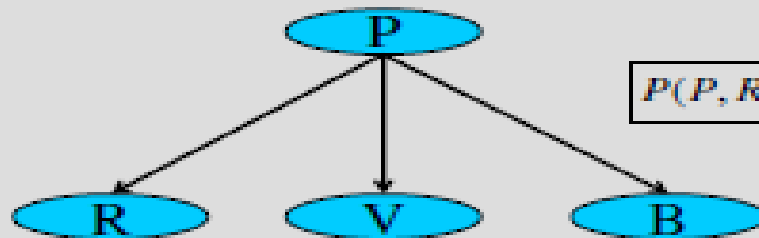
Exemple d'application d'un réseau bayésien Naïf (2/5)

Définition d'une image

- Une image (format jpeg ou gif) est un tableau de points;
- Chaque point est défini par l'intensité/nuance de 3 composantes: rouge, vert & bleu;
- Chaque intensité/nuance varie de 0 à 255;
- Par exemple :
 - Un point dont les composantes RVB sont (0, 0, 0) est de couleur noire;
 - Un point dont les composantes RVB sont (255, 255, 255) est de couleur blanche;
 - Un point dont les composantes RVB sont (255, 0, 0) est de couleur rouge.

Exemple d'application d'un réseau bayésien Naïf (3/5)

Réseau bayésien naïf modélisant le problème



$$P(P, R, V, B) = P(P) \times P(R|P) \times P(V|P) \times P(B|P)$$

- P : le point appartient au visage ou pas (variable booléenne);
- On va travailler avec des composantes RVB normalisées;

$$R = \frac{R}{R + V + B}$$

$$V = \frac{V}{R + V + B}$$

$$B = \frac{B}{R + V + B}$$

- R : le pourcentage de rouge dans le point (variable entière entre 0 et 100);
- V : le pourcentage de vert dans le point (variable entière entre 0 et 100);
- B : le pourcentage de bleu dans le point (variable entière entre 0 et 100);

Exemple d'application d'un réseau bayésien Naïf (4/5)

Apprentissage du réseau bayésien naïf(1/2)

- On dispose d'un ensemble d'images pour lesquels on connaît les points qui appartiennent au visage et ceux qui n'y appartiennent pas, ainsi que l'intensité des composantes RVB : images étiquetées;
- Pour chaque point, on dispose d'un quadruplet (P, R, V, B);



Olivier.aycard@imag.fr

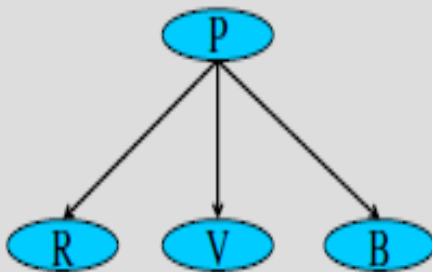


UNIVERSITÉ JOSEPH FOURIER
1385 AVENUE DENIS POISSON
38000 GRENoble, FRANCE

32

Exemple d'application d'un réseau bayésien Naïf (5/5)

Apprentissage du réseau bayésien naïf(2/2)



Il suffit de compter

P	R	V	B
1	250	10	20
1	248	12	25
0	150	50	12
0	23	50	89
0	56	50	46
0	80	70	35
0	5	56	98
0	200	140	233
0	200	235	57
0	150	56	34

$$P(P = \text{vrai}) = 2/10$$

$$P(P = \text{faux}) = 8/10$$

$$P(R = 250 | P = \text{vrai}) = 1/2$$

$$P(R = 248 | P = \text{vrai}) = 1/2$$

$$P(V = 50 | P = \text{faux}) = 3/8$$

Utilisation du réseau bayésien naïf pour la reconnaissance d'une personne

Pour chaque point d'une nouvelle image, on calcule sa probabilité d'appartenir au visage ou pas;

$$P(P = \text{vrai} | R = r, V = v, B = b) = \alpha \times P(P = \text{vrai}) \times P(R = r | P = \text{vrai}) \times P(V = v | P = \text{vrai}) \times P(B = b | P = \text{vrai})$$

$$P(P = \text{faux} | R = r, V = v, B = b) = \alpha \times P(P = \text{faux}) \times P(R = r | P = \text{faux}) \times P(V = v | P = \text{faux}) \times P(B = b | P = \text{faux})$$

On détermine ensuite α et on normalise;

On calcule la position du visage dans l'image;

$$X = \frac{\sum_{i,j} P(P_{i,j} = \text{vrai} | R_{i,j} = r, V_{i,j} = v, B_{i,j} = b) \times x}{\sum_{i,j} P(P_{i,j} = \text{vrai} | R_{i,j} = r, V_{i,j} = v, B_{i,j} = b)}$$

$$Y = \frac{\sum_{i,j} P(P_{i,j} = \text{vrai} | R_{i,j} = r, V_{i,j} = v, B_{i,j} = b) \times y}{\sum_{i,j} P(P_{i,j} = \text{vrai} | R_{i,j} = r, V_{i,j} = v, B_{i,j} = b)}$$

Conclusion

+ méthode très répandue

+ facile à implanter

+ facilite la recherche

+ simplifie la mise à jour

**+ utilise un certain nombre d'avantages des réseaux bayésiens :
estimation des données incomplètes.**

Exercices (TD)