Réseaux Bayésiens

Bruno Bouzy

26 février 2008

Introduction

Ce chapitre présente les réseaux bayésiens à partir du tutoriel d'Andrew Moore (http://www.autonlab.org/tutorials/bayesnet.html). Il discute de l'utilisation des probabilités jointes pour décrire l'incertitude de faits. Il montre comment les Réseaux Bayésiens (RB) permettent de construire ces probabilités jointes, et comment à partir de ces probabilités jointes, on peut retrouver toutes les probabilités conditionnelles souhaitées. Selon Andrew Moore, les réseaux bayésiens constituent la technologie la plus puissante de ces 10 dernières années en IA et en apprentissage automatique. Les RB constituent un langage graphique et une méthodologie, simples et corrects, pour exprimer pratiquement ce de quoi on est certain ou incertain. Ils reposent sur la formule de Bayes reliant des probabilités conditionnelles avec des probabilités jointes.

Formule de Bayes

Il y a des faits, désignés par A ou B, qui ont des probabilités d'arriver P(A) et P(B). P(~A) est la probabilité que non A arrive. La formule de Bayes dit que :

$$P(A|B) = P(A, B)/ P(B) \qquad (1)$$

ou encore:

$$P(A|B) = P(B|A).P(A)/ P(B) \qquad (2)$$

Etant donné que :

$$P(B) = P(B|A).P(A) + P(B|{\sim}A).P({\sim}A) \qquad (3)$$

On écrit alors:

$$P(A|B) = P(B|A).P(A)/(P(B|A).P(A) + P(B|{\sim}A).P({\sim}A)) \qquad (4)$$

La formule de Bayes peut être conditionnée par un fait X:

$$P(A|B,X) = P(B|A,X).P(A,X)/ P(B,X) \qquad (5)$$

Une évidence utile à préciser:

$$P(A|B) + P({\sim}A|B) = 1 \qquad (6)$$

Une table de probabilités jointes

Soient 3 variables A, B, C pouvant valoir vrai ou faux. On peut écrire une table listant toutes les combinaisons de ces 3 variables. Il y a 2^3 combinaisons. Pour chacune de ces combinaisons, on peut donner la probabilité jointe de la combinaison. La table constitue donc la distribution de probabilité jointe des variables A, B, C. La somme des probabilités dans la table vaut 1.

Si on cherche la probabilité que A soit vrai, il suffit de sommer toutes les probabilités des combinaisons de la table pour lesquelles A est vrai.

Si on cherche la probabilité que A soit vrai et B vrai, il suffit de sommer toutes les probabilités des combinaisons de la table pour lesquelles A est vrai et B vrai. Elémentaire.

Si on cherche la probabilité que A soit vrai **sachant que** B est vrai, d'après (1), il suffit de sommer toutes les probabilités des combinaisons de la table pour lesquelles A est vrai et B est vrai et de diviser par la somme des probabilités des combinaisons de la table pour lesquelles B est vrai.

La bonne nouvelle: Si on a la distribution de probabilité jointe dans une table, on peut en déduire n'importe quelle probabilité conditionnelle.

La mauvaise nouvelle: Dés que le nombre de variables augmente, la table devient impossible à construire car sa taille est exponentielle en fonction du nombre de variables.

Dans la suite, nous allons voir comment l'indépendance entre certaines variables permet de remplacer la table de probabilités jointes par un réseau bayésien dont la taille est largement inférieure à celle de la table. Pour cela, sur des exemples simples, nous allons voir les deux cas d'indépendance: indépendance tout court et indépendance conditionnelle. Cela nous permettra d'introduire l'intérêt des RB et le formalisme graphique associé. Nous verrons ensuite comment utiliser un RB pour calculer une probabilité jointe ou une probabilité conditionnelle quelconque.

Un cas d'indépendance

Supposons que l'on ait deux variables:

M: Manuela enseigne (sinon c'est Andrew),
S: il fait beau (Sunny).

La table de probabilités jointes contient 4 combinaisons. Est-ce que nous devons vraiment calculer les 4 probabilités jointes correspondant à ces 4 combinaisons ?

Dans cet exemple, le fait qu'il fasse beau n'influence pas le fait que Manuela enseigne et n'en dépend pas. Cela peut s'écrire:

$$P(S|M) = P(S) \qquad (7)$$

Pour écrire (7), on a donc utilisé une connaissance du domaine. L'équation (7) revient à écrire:

$$P(\sim S|M) = P(\sim S) \qquad (8)$$
$$P(M|S) = P(M) \qquad (9)$$
$$P(M, S) = P(M)P(S) \qquad (10)$$
$$P(M, \sim S) = P(M)P(\sim S) \qquad (11)$$
$$P(\sim M, S) = P(\sim M)P(S) \qquad (12)$$
$$P(\sim M, \sim S) = P(\sim M)P(\sim S) \qquad (13)$$

Donc si on connait P(S) et P(M):

$$P(S) = 0.3 \qquad (14)$$
$$P(M) = 0.6 \qquad (15)$$

alors on est capable de remplir complètement la table des probabilités jointes avec ces 2 nombres. L'indépendance entre M et S permet de construire la table avec **2 nombres** au lieu de **3 nombres**. (Avec 3 probabilités jointes, on trouve la 4ème par différence à 1).

Si on rajoute un fait:

L: L'enseignant arrive en retard (Late).

Et si on suppose que l'enseignant peut être retardé par le mauvais temps et que Andrew arrive plus souvent en retard que Manuela.

On sait donc que L dépend de S et M. Supposons que l'on connaisse cette dépendance sous forme de probabilités conditionnelles. Par exemple:

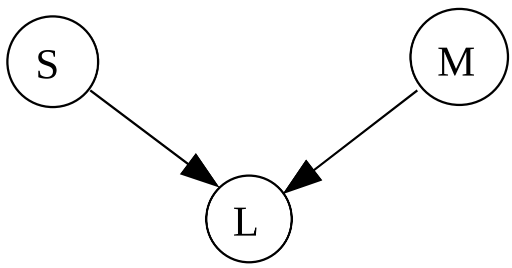
$$P(L \mid M, S) = 0.05 \qquad (16)$$
$$P(L \mid M, \sim S) = 0.1 \qquad (17)$$
$$P(L \mid \sim M, S) = 0.1 \qquad (18)$$
$$P(L \mid \sim M, \sim S) = 0.2 \qquad (19)$$

On s'aperçoit alors que l'on peut trouver **toutes** les probabilités jointes avec les **6 nombres** P(S), P(M), P(L | M, S), P(L | M, ~S), P(L | ~M, S) et P(L | ~M, ~S), alors que sans l'indépendance entre M et S, il faudrait **7 nombres** (8 – 1).

Finalement, sur ce petit exemple avec 3 variables dont 2 sont indépendantes, on voit que l'on peut avoir 6 nombres au lieu de 7 pour connaitre toutes les probabilités jointes. Avec un nombre plus grand de variables, dont certaines sont indépendantes, les gains en taille de la table de probabilités jointes seront encore plus grands.

Notation graphique

A chaque variable, on associe un noeud. Et à chaque relation de dépendance entre deux variables, on associe une flèche reliant les deux noeuds associés.



De plus, à chaque noeud on associe les équations permettant de calculer les probabilités jointes. Au noeud S, on ajoute l'équation (14). Au noeud M, on ajoute l'équation (15). Au noeud L, on ajoute les équations (16), (17), (18) et (19).

Indépendance conditionnelle

Supposons que l'on ait 3 variables:

M: Manuela enseigne (sinon c'est Andrew),

L: L'enseignant arrive en retard (Late).

R: Le sujet de l'enseignement concerne les Robots.

Et que:

Andrew a plus de chance que Manuela d'arriver en retard.
Andrew a plus de chance que Manuela d'enseigner un cours sur les robots.

Quelle sorte d'indépendance peut on identifier ?

L dépend de M
R dépend de M
L et R ne dépendent pas l'un de l'autre.

Que peut-on dire des égalités suivantes ?

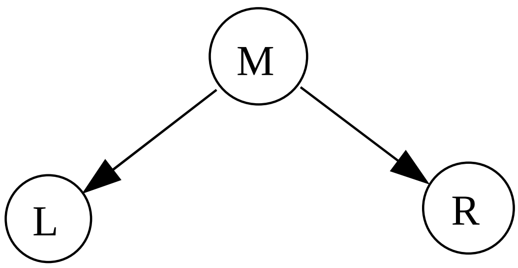
$P(L|M) = P(L)$ non c'est faux car L dépend de M.
$P(R|M) = P(R)$ non c'est faux car R dépend de M.
$P(L|R) = P(L)$ oui c'est vrai car L et R ne dépendent pas l'un de l'autre.

Une fois que l'on sait qui enseigne (M est fixé), alors le fait de savoir si l'enseignant arrive en retard ou pas n'influence pas sur le contenu du cours (sur les robots ou pas). Ceci s'exprime par:

$P(R|M,L) = P(R|M)$ et $P(R|\sim M,L) = P(R|\sim M)$

Et cela se dit: R et L sont indépendants étant donné M.

Graphiquement, cela donne:



Plus formellement, R et L sont indépendants étant donné M si:

Pour tout x, y, z dans {vrai, faux} on a $P(R=x \mid M=y, L=z) = P(R=x \mid M=y)$

Pour connaître toutes les probabilités jointes, il suffit de connaître $P(M)$, $P(L \mid M)$, $P(L \mid \sim M)$, $P(R \mid M)$ et $P(R \mid \sim M)$, c'est-à-dire **5 nombres** (au lieu de 7 nombres sans l'indépendance conditionnelle).

$$P(L \mid M) = 0.085 \qquad (20)$$
$$P(L \mid \sim M) = 0.17 \qquad (21)$$
$$P(R \mid M) = 0.3 \qquad (22)$$
$$P(R \mid \sim M) = 0.6 \qquad (23)$$

Un exemple avec 5 variables

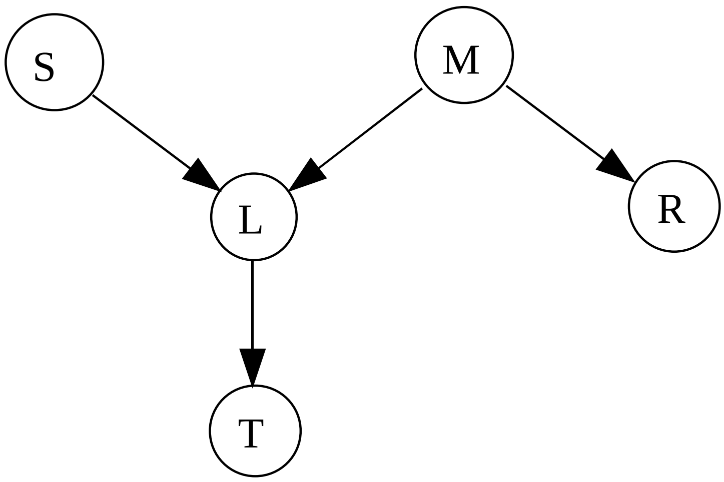
Supposons que l'on ait 5 variables:

T:	Le cours commence à 10:35.
L:	L'enseignant arrive en retard (Late).
R:	Le sujet de l'enseignement concerne les Robots.
M:	Manuela enseigne (sinon c'est Andrew).
S:	il fait beau (Sunny).

avec les indépendances conditionnelles suivantes:

T est influencé directement par L, mais est indépendant de R, M, S connaissant L.
L est influencé directement par M et S, mais est indépendant de R, connaissant M et S.
R est influencé directement par M, mais est indépendant de R, M, S connaissant L.
M et S sont indépendants.

Graphiquement, cela donne:



Pour connaître toutes les probabilités jointes il suffit de connaitre P(S), P(M), P(L | M, S), P(L | M, ~S), P(L | ~M, S), P(L | ~M, ~S), P(R | M), P(R | ~M), P(T | L) et P(T | ~L), c'est-à-dire **10 nombres** seulement. Sans les indépendances, il faudrait 32 – 1 = **31 nombres**.

$$P(T \mid L) = 0.3 \qquad (24)$$
$$P(T \mid \sim L) = 0.8 \qquad (25)$$

Réseaux Bayésiens formalisés

Un réseau bayésien est un graphe orienté acyclique avec un ensemble N de noeuds et un ensemble A d'arcs orientés.

Un noeud contient:
+ le nom d'une variable aléatoire,
+ une table de probabilités indiquant comment la probabilité de cette variable dépend des combinaisons de valeurs parentales.

Pour construire un Réseau Bayésien:

1. Choisir un ensemble variables pertinentes ordonnées $X_1, X_2, ..., X_m$.
2. Pour i=1 à m
    1. Ajouter X_i au graphe
    2. Parents(X_i) = sous-ensemble minimal de $\{X_1, ..., X_{i-1}\}$ tel qu'il y ait indépendance conditionnelle de X_i et de tous les autres éléments de $\{X_1, ..., X_{i-1}\}$ étant donné Parents(X_i)
    3. Définir la table de probabilités $P(X_i=k|\text{valeurs affectées à Parents}(X_i)$

Calcul d'une probabilité jointe avec un RB

Exemple: comment calculer P(S, ~M, L, ~R, T) ?

On applique successivement la formule de Bayes avec une simplification d'indépendance:

On choisit la variable T:

$$\begin{aligned} P(S, \sim M, L, \sim R, T) &= P(T \mid S, \sim M, L, \sim R)\ \ P(S, \sim M, L, \sim R) \\ &= P(T \mid L)\ \ P(S, \sim M, L, \sim R) \end{aligned}$$

car T ne dépend que de L. On recommence avec la variable R:

$$\begin{aligned} P(S, \sim M, L, \sim R) &= P(\sim R \mid S, \sim M, L)\ \ P(S, \sim M, L) \\ &= P(\sim R \mid \sim M)\ \ P(S, \sim M, L) \end{aligned}$$

car R ne dépend que de M. On choisit la variable L:

$$P(S, \sim M, L) = P(L \mid S, \sim M)\ \ P(S)\ \ P(\sim M)$$

Finalement:

$$P(S, \sim M, L, \sim R, T) = P(T \mid L)\ \ P(\sim R \mid \sim M)\ \ P(L \mid S, \sim M)\ \ P(S)\ \ P(\sim M) \qquad (26)$$

On trouve les valeurs numériques des ces probabilités dans le RB:

$$\begin{aligned} P(S, \sim M, L, \sim R, T) &= 0.3\ (1-0.6)\ 0.1\ 0.3\ (1-0.6) = 0.12\ 0.1\ 0.12 \\ &= 0.00144 \end{aligned}$$

Dans le cas général, pour calculer une probabilité jointe, on a:

$$P(X_1=x_1, X_2=x_2, ..., X_n=x_n) = \prod_i P(X_i=x_i|\text{affectations de Parents}(X_i)) \qquad (27)$$

Calcul d'une probabilité conditionnelle arbitraire avec un RB

Pour calculer une probabilité conditionnelle arbitraire (c'est-à-dire n'étant pas explicitement dans le RB), on se ramène à des probabilités jointes (que l'on sait calculer).

On écrit par exemple:

$$P(R \mid T, \sim S) = P(R, T, \sim S) / (P(R, T, \sim S) + P(\sim R, T, \sim S)) \qquad (28)$$

Bilan

Les bonnes nouvelles:

On dispose d'une méthode pour construire un RB.
On évite le stockage exponentiel en fonction du nombre de variables.
On a un stockage exponentiel en fonction du nombre maximum de parents des noeuds.
On est capable de calculer n'importe quelle probabilité jointe, dans un temps proportionnel au nombre de noeuds.
On est capable de calculer n'importe quelle probabilité conditionnelle.

La mauvaise nouvelle:

Ce dernier calcul est exponentiel en fonction du nombre de variables.

Méthodes d'échantillonnage

Pour calculer une probabilité conditionnelle et éviter le calcul exponentiel en fonction du nombre de variables, on peut faire une simulation stochastique (stochastique = contenant du hasard).

Sur notre exemple, on veut calculer $P(R \mid T, \sim S)$.

On effectue N fois:

1. Choisir au hasard S avec S=vrai avec probabilité 0.3.
2. Choisir au hasard M avec M=vrai avec probabilité 0.6.
3. Choisir au hasard L avec L=vrai respectant la table du noeud L.
4. Choisir au hasard R avec R=vrai respectant la table du noeud R.
5. Choisir au hasard T avec T=vrai respectant la table du noeud T.
6. si T=vrai et S=faux alors Nc++
7. si R=vrai et T=vrai et S=faux alors Ns++

A la fin:

Nc/N est une estimation de $P(T, \sim S)$.
Ns/N est une estimation de $P(R, T, \sim S)$.
donc Ns/Nc est une estimation de $P(R \mid T, \sim S)$.

Ce résultat particulier à l'exemple se généralise au calcul de P(X | Y).

Le problème de la simulation stochastique est que beaucoup d'échantillons sont rejetés. Dans notre exemple, tous les échantillons avec S=vrai sont rejetés, c'est-à-dire 30% des échantillons. L'idée est donc d'obliger le tirage S=faux et de pondérer le résultat pas la probabilité de P(~S) = 0.7.

En général, si à la fin, on ne compte que les échantillons tels que Xi=v, et si l'échantillon tiré au hasard au début est tel que Xi!=v, alors cela ne sert à rien de continuer. Donc, au lieu de faire cela, on oblige le tirage de l'échantillon avec Xi=v, et on normalise en multipliant par p, la probabilité que Xi=v.

Cela donne l'algorithme de simulation stochastique avec poids, suivant:

Soit à calculer $P(E_1|E_2)$ où E_1 et E_2 sont des ensembles d'affectations de variables.
Nc = 0 et Ns = 0
faire N fois:

1. Engendrer une affectation aléatoire des variables matchant E_2
2. Mesurer la probabilité de cette affectation aléatoire W
3. Nc = Nc + W
4. si matching de E_1 alors Ns = Ns + W

Finalement Ns/Nc est une estimation de $P(E_1|E_2)$.

A retenir

La signification et l'importance de l'indépendance et de l'indépendance conditionnelle.

La définition d'un réseau bayésien.

Le calcul des probabilités d'affectation des variables (i.e. Calcul d'une probabilité jointe) dans un réseau bayésien.

La méthode lente (exponentielle) pour calculer une probabilité conditionnelle arbitraire.

La méthode stochastique pour calculer une probabilité conditionnelle arbitraire.

Références

Andrew W. Moore, Bayesian Networks, Tutorial Slides (http://www.autonlab.org/tutorials/bayesnet.html), 2004.

Antoine Cornuéjols, Laurent Miclet, Apprentissage artificiel, concepts et algorithmes, Eyrolles, pages 57-64.

Christopher Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995, chapitre 1, pages 17-28.

Exercice

Une alarme est située dans une maison. Normalement, l'alarme sonne (A) lorsqu'un voleur cambriole la maison (V). On a P(V) = 0.01. On connaît les probabilités conditionnelles P(A|V) et P(A|~V).

Question 1

Construire le réseau bayésien.

Une alarme est «complète» ssi elle détecte tous les voleurs: P(A|V)=1. Une alarme est «correcte» ssi elle ne sonne que si un voleur est présent: P(A|~V)=0.

L'erreur de complétude est égale à 1-P(A|V). L'erreur de correction est égale à P(A|~V). L'erreur totale est la somme des deux.

Question 2

Pour chacun des cas suivants:

P(A|V) = 1	P(A|~V) = 0.001
P(A|V) = 1	P(A|~V) = 0.1
P(A|V) = 0.999	P(A|~V) = 0
P(A|V) = 0.9	P(A|~V) = 0

a) Calculer les probabilités jointes P(A,V), P(A,~V), P(~A,V), P(~A,~V).

b) En déduire P(V|A) et P(V|~A) et l'erreur totale pour chacun des cas.

c) Conclure sur l'importance relative de la correction et complétude des alarmes.