DIC9150 Concepts fondamentaux de l'informatique cognitive

Réseaux Bayésiens

Roger Villemaire

Département d'informatique
UQAM

14 novembre 2023

Plan

Conviction (Belief)

- Contrairement au point de vue idéalisé de la logique (classique) où les connaissances sont certaines (vraies ou fausses), la connaissance est souvent incertaine, car :
  - on ignore des faits, on manque de données (Laplace),
  - on n'a pas de principes explicatifs qui font consensus,
  - le principe explicatif est justement que l'incertitude ne peut pas être éliminée (mécanique quantique).

Hasard

- Que sont le hasard, la chance, l'aléatoire ?
- Pascal répond à la question du partage équitable des paris :
  - qu'ils doivent être faits en proportion des différentes possibilités.
- Point de départ de l'analyse mathématique du hasard.
  - Un événement aléatoire est déterminé par sa probabilité qui est un nombre réel entre 0 et 1.
  - Intuitivement une probabilité de $\alpha \in [0, 1]$ signifie que sur un grand nombre d'essais N, l'événement devrait se produire $\alpha \times N$ fois.

Exemple

- $V \in \{1, 2, 3, 4, 5, 6\}$ est la valeur que peut prendre un dé.
  - On dit que V est une *variable aléatoire*.
  - Comme elle ne peut prendre qu'un nombre fini (donc énumérable) de valeurs, il s'agit d'une variable *discrète* (sinon c'est une valeur dans les réels et elle est dite *continue*).
- La *distribution* de probabilité $P(V)$ de la variable V est une fonction qui associe à chaque valeur de V une probabilité de telle façon à ce que la somme de ces valeurs soit 1.
  - $P(V = i) = 1/6 = 0{,}17$ pour $i = 1, 2, 3, 4, 5, 6$ (équiprobable) pour un dé non pipé.

Prise de décision

- Les probabilités servent à la prise de décision :
  - le partage des paris de Pascal !
  - théorie économique (maximiser l'espérance du gain),
  - systèmes d'aide à la décision, systèmes experts, robots etc.
  - voir l'exposé de J. Halpern de l'école d'été de l'ISC 2016.
- Néanmoins il est très bien documenté que les humains ont tendance à mésestimer les probabilités et les risques :
  - vaste littérature en psychologie, voir les premiers jours de l'école d'été ISC 2016 !
- Il est d'autant plus utile d'avoir des outils informatiques pour faire, évaluer et valider ces calculs.

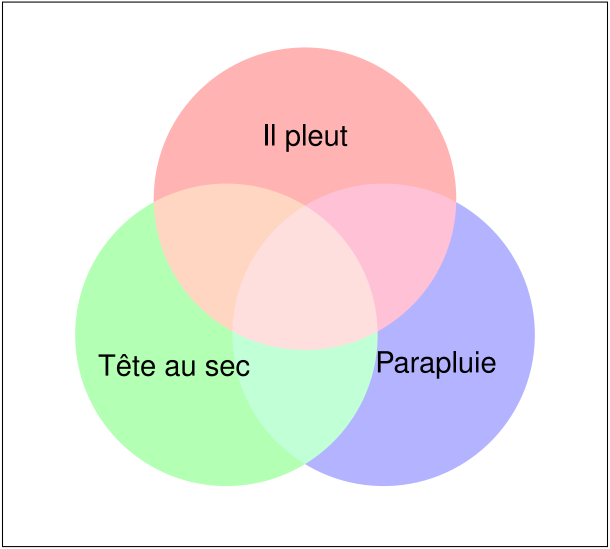
Terminologie

- Une variable aléatoire (discrète) V de domaine dom_V est une fonction qui associe une probabilité (valeur réelle entre 0 et 1) à chaque élément de dom_V de telle façon à ce que le total soit de 1.
- Un événement aléatoire est une condition sur des variables aléatoires, comme par exemple :
  - $V = d$, pour un $d \in dom_V$,
  - $V \neq d \cup V \neq d'$, pour des $d, d' \in dom_V$.

Calcul des probabilités

- Pour saisir les probabilités, en les visualisant, il est très utile de s'aider à l'aide de diagrammes de Venn.
- Les événements aléatoires correspondent alors à des unions des "zones" du diagramme de Venn.
- La probabilité $P(E)$ d'un événement E est une notion de "taille", un réel.
- La probabilité totale de tout l'espace est $P(\mathit{True}) = 1$.

Exemple



Probabilité *a priori*

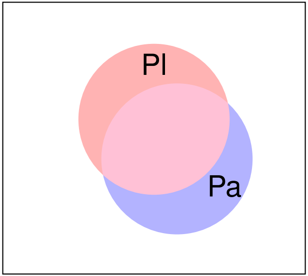
- Avec E et E' des événements aléatoires, on a
  - $P(E)$, la probabilité de E,
  - $P(E')$, la probabilité de E'.
- Ce sont des mesures de la “chance” que ces événements surviennent *étant donné que l’on ne sait rien sur les événements qui se sont déjà produits*.
  - Cas extrême : Si E' s’est produit, on en est sûr et donc sa probabilité *a priori* n’est pas un bon indicateur !

Probabilité *a posteriori*

- Si l'on sait que E' s'est produit, le bon indicateur de la "chance" que E se produise maintenant est
  - $P(E|E')$, la probabilité de E *conditionnelle* à E' (la probabilité de E *étant donné* E').
- Définition : $P(E|E') = P(E \cap E')/P(E')$.

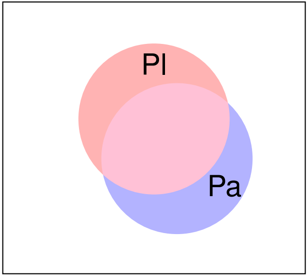
Exemple

- Pl : il pleut,
- Pa : j'ai un parapluie avec moi,
- $P(Pa)$ la probabilité que j'ai un parapluie avec moi,
- $P(Pa|Pl)$ la probabilité que j'ai un parapluie avec moi, étant donné qu'il pleut,
- $P(Pa|Pl) = P(Pa \cap Pl)/P(Pl)$.



Exemple, calculs

- $P(Pl) = 0{,}20$
- $P(Pa) = 0{,}22$
- $P(Pa \cap Pl) = 0{,}15$
- $P(Pa|Pl) = \frac{P(Pa \cap Pl)}{P(Pl)} = \frac{0{,}15}{0{,}20} = 0{,}75$



Indépendance conditionnelle

- A et B sont conditionnellement indépendants étant donné C si
  - $P(A|B,C) = P(A|C)$,
- Intuitivement :
  - étant donné C, B n'a pas d'influence sur (la probabilité d'avoir) A,
  - la probabilité que A se produise étant donné C n'est pas changée si l'on sait, de plus, que B s'est produit.
  - ceci est le cas si le fait d'avoir B n'influence pas le fait d'avoir A, mais pas réciproquement !

Réseaux Bayésiens

- Un *réseau Bayésien* (Bayesian ou Belief Network) est
  - un graphe orienté acyclique dont les noeuds sont des variables aléatoires,
  - tel que pour tout noeud V, un noeud W qui n'est pas un descendant de V est conditionnellement indépendant de V étant donné $parents(V)$
    - $parents(V) = \{V'; V' \to V \text{ dans le réseau}\}$
- Intuitivement la probabilité d'avoir V étant donnés ses $parents(V)$, n'est pas modifiée par l'ajout d'un non-descendant W.

Réseaux Bayésiens

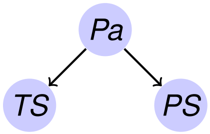
- Un réseau bayésien représente donc les dépendances conditionnelles.
- Il est important de noter que les parents d'un noeud ne sont pas nécessairement des causes, mais les causes permettent en général des réseaux bayésiens compact.
  - En fait, on n'a que le relation :
    - $P(\text{noeud}|\text{parents et des non-descendants}) = P(\text{noeud}|\text{parents})$

Probabilité totale

- Un réseau Bayésien est une description, en général compacte, de la *distribution de probabilité totale* (joint probability distribution) de tous ses noeuds.
- Pour $V_1, \ldots, V_n$ les variables aléatoires du réseau, la distribution de probabilité totale $P(V_1, \ldots, V_n)$ détermine la probabilité de toutes les combinaisons de valeurs des variables.
  - Elle permet donc de déterminer la probabilité de n'importe quel événement aléatoire défini à partir de ces variables.
  - C'est donc un description complète du système probabiliste.

Exemple

- $P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$ donc
  $P(X,Y) = P(X \cap Y) = P(X|Y) \cdot P(Y)$
- $P(TS, PS, Pa) =$
  $P(TS|PS, Pa) \cdot P(PS, Pa) =$
  $P(TS|PS, Pa) \cdot P(PS|Pa) \cdot P(Pa) =$
  $P(TS|Pa) \cdot P(PS|Pa) \cdot P(Pa)$



Usages

- Très souvent un expert du domaine pourra déterminer correctement les relations causales entre les variables et ainsi établir un réseau bayésien correct.
- Néanmoins, l'expérience et les recherches en psychologie ont montrés qu'il était beaucoup plus difficile d'établir correctement les probabilités.
- Il est néanmoins possible d'automatiser ce processus en récupérant des données, pour calculer les fréquences et s'en servir comme valeurs des probabilités.

Variables latentes

- Néanmoins certaines variables peuvent être difficiles voire impossibles à mesurer directement, il s'agit de variables *latentes*.
- Il est alors possible, avec des méthodes analogues à celles utilisées avec les réseaux neuronaux de déterminer des probabilités pour les variables qui minimisent les différences entre les valeurs mesurées (les données) et celles prédites par le réseaux.

Autres applications

- Lorsque le réseau est établi, avec ses probabilités, on peut faire générer des données pseudo-aléatoires satisfaisant ces distributions.
- Il est alors possible de calculer les fréquences pour n'importe quelle événement. Il s'agit donc d'une méthode de calcul par simulation.
- Le domaine est vaste et on peut mentionner pour terminer qu'il existe aussi des méthodes pour
  - déterminer ou transformer un réseau bayésien à partir de données,
  - comparer les réseaux bayésiens et ainsi valider le modèle.

Conclusion

- L'incertitude est très présente dans les applications,
- Les probabilités sont une méthode de représentation de l'incertitude ayant fait ses preuves (méthodes de calcul bien établies).
- Les réseaux bayésiens permettent en général une représentation compacte, permettant des calculs assez efficaces pour la fonction de probabilité totale.
- Ceci nous donne une infrastructure permettant d'effectuer les calculs nécessaires pour répondre à essentiellement n'importe quelle question de nature probabiliste au sujet des variables aléatoires du réseau.