INTRODUCTION

La modélisation probabiliste de problèmes complexes passe systématiquement par la définition d'une loi jointe multivariée pour l'ensemble des variables considérées. Si on note $(X_i)_{i=1\ldots n}$ un ensemble de n variables aléatoires réelles, leur loi jointe s'écrit :

$$\mathbb{P}(X_1, X_2, \ldots, X_n) = \mathbb{P}(X_1) \times \mathbb{P}(X_2|X_1) \times \ldots \times \mathbb{P}(X_n|X_1, \ldots, X_{n-1})$$

En admettant que ces variables prennent un nombre k ($k \geqslant 2$) de valeurs, la probabilité précédente doit être définie un ensemble de k^n configurations, c'est-à-dire qu'il faut définir un total de $k^n - 1 = (k-1) + k(k-1) + \ldots k^{n-1}(k-1)$ paramètres libres.

Même dans le cas de variables binaires ($k = 2$) et même pour un faible nombre de variables (ex: $n = 100$), le nombre de paramètres à définir devient rapidement gigantesque (ex: $2^{100} - 1 \simeq 10^{30}$), d'où la nécessité absolue de développer des modèles plus parcimonieux.

Dans ce but, on réécrit la loi jointe sous la forme :

$$\mathbb{P}(X_1, X_2, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{n} \mathbb{P}(X_i|\mathrm{pa}(X_i))$$

où $\mathrm{pa}(X_i)$ est l'ensemble quelconque des variables "parentes" de X_i et où $Z > 0$ est une constante de normalisation. Cette factorisation définit la classe très large des *modèles graphiques probabilistes* car la fonction $\mathrm{pa}()$ relie les variables sous la forme d'un graphe.

Dans le cas où $\mathrm{pa}(X_i) \subset \{X_1, \ldots, X_{i-1}\}$ pour tout i (pour un ordre donné), le graphe obtenu a la propriété d'être un *graphe orienté sans circuit*, $Z = 1$, et ces modèles graphiques s'appellent *réseaux bayésiens*[1]. Dans le cas où la relation parentale est symétrique ($Y \in \mathrm{pa}(X) \iff X \in \mathrm{pa}(Y)$) le graphe peut être interprété comme non orienté les modèles résultants s'appellent des *champs markoviens aléatoires* où *réseaux markoviens* (en général, $Z \neq 1$ pour ces modèles).

Ces modèles graphiques probabilistes sont en fait omniprésents dans notre vie de tous les jours. On peut, par exemple, penser à la reconnaissance vocale de nos téléphones portables qui utilise des *chaînes de Markov cachées* (cas particulier des

1. Attention, s'il est évidemment possible de faire de l'inférence bayésienne avec des réseaux bayésiens, le terme 'bayésien' est ici employé pour faire référence à la définition du modèle à l'aide de lois conditionnelles.

réseaux bayésiens). Ces modèles peuvent aussi être présents dans le domaine de la génétique, que ce soit les consultations prénatales ou celles qui concernent la génétique du cancer, dans les deux cas les modèles sous-jacents font de manière quasi systématique intervenir des réseaux bayésiens. On peut également penser à la reconnaissance de formes et/ou à la segmentation en imagerie (par exemple, médicale ou télésurveillance) qui s'appuie pour l'essentiel sur des réseaux markoviens. Les domaines d'application de ces modèles sont extrêmement variés : acoustique, biologie des systèmes, réseaux sociaux, transport urbain, contrôle qualité, etc. Les modèles graphiques probabilistes et leurs variantes interviennent en fait dans un nombre considérable de problèmes en lien avec l'intelligence artificielle : classification supervisée ou non, théorie de la décision, systèmes experts, causalité, etc.

Dans ce numéro spécial, nous présentons une sélection d'articles de recherche utilisant largement les modèles graphiques probabilistes et initialement présentés dans une version courte lors des 7e Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes (JFRB) organisées à Paris en juin 2014. Le premier article s'intéresse au problème difficile de l'apprentissage de la structure d'un modèle graphique probabiliste (ici, non orienté) et se propose de l'approcher avec une méthode bayésienne innovante se focalisant sur une classe très particulière de réseaux : les arbres. Après cette première contribution méthodologique, le deuxième article propose d'utiliser ces outils pour modéliser le comportement d'un ensemble d'agents, avec une application à l'étude des trajectoires dans une colonie d'abeilles. Le troisième article évoque l'épineux problème de la recherche de causalité dans un réseau de gènes et propose une méthode originale permettant de combiner à la fois des données d'observation et d'intervention. Le dernier article, enfin, propose une ouverture vers les approches possibilistes dont l'objectif est de présenter une alternative aux approches probabilistes afin de mieux prendre en compte les informations approximatives ou imprécises auxquelles nous sommes très souvent confrontés dans les applications réelles.

Nous remercions tous les relecteurs qui ont contribué à l'élaboration de ce numéro en veillant sur la qualité de son contenu.

PHILIPPE LERAY
Université de Nantes
LINA UMR CNRS 6241
Equipe DUKe (Data User Knowledge)
Nantes, France

GRÉGORY NUEL
Institut des mathématiques, CNRS
LPMA, UPMC
Paris, France

COMITÉ DE LECTURE