GNU LINUX MAGAZINE / FRANCE

Ceci est un extrait électronique d'une publication de
Diamond Editions :

http://www.ed-diamond.com

Retrouvez sur le site tous les anciens numéros en vente par
correspondance ainsi que les tarifs d'abonnement.

Pour vous tenir au courant de l'actualité du magazine, visitez :

http://www.gnulinuxmag.com

Ainsi que :

http://www.linux-pratique.com

et

http://www.miscmag.com
GNU LINUX MAGAZINE / FRANCE

Ceci est un extrait électronique d'une publication de Diamond Editions

http://www.ed-diamond.com

Creative Commons

Paternité - Pas d'Utilisation Commerciale - Pas de Modification
2.0 France

Vous êtes libres :

- de reproduire, distribuer et communiquer cette création au public.

Paternité. Vous devez citer le nom de l'auteur original de la manière indiquée par l'auteur de l'oeuvre ou le titulaire des droits qui vous confère cette autorisation (mais pas d'une manière qui suggérerait qu'ils vous soutiennent ou approuvent votre utilisation de l'oeuvre).

Pas d'Utilisation Commerciale. Vous n'avez pas le droit d'utiliser cette création à des fins commerciales.

Pas de Modification. Vous n'avez pas le droit de modifier, de transformer ou d'adapter cette création.

A chaque réutilisation ou distribution de cette création, vous devez faire apparaître clairement au public les conditions contractuelles de sa mise à disposition.

- Chacune de ces conditions peut être levée si vous obtenez l'autorisation du titulaire des droits.
- Rien dans ce contrat ne diminue ou ne restreint le droit moral de l'auteur ou des auteurs.

Ceci est le Résumé Explicatif du Code Juridique. La version intégrale du contrat est attachée en fin de document et disponible sur :

http://creativecommons.org/licenses/by-nc-nd/2.0/fr/legalcode

Par :: Jean-Michel Marin :: marin@ceremade.dauphine.fr ::
:: Fabrice Rossi :: Fabrice.Rossi@apiacoa.org ::

Découvrez les réseaux Bayésiens

Le terme « Bayésien » est très en vogue depuis l'apparition des filtres de spam ainsi qualifiés. Nous nous proposons dans cet article de le démythifier en présentant un modèle mathématique extrêmement puissant, les réseaux Bayésiens. Nous verrons comment ces outils généralisent la notion de raisonnement classique en le plongeant dans un cadre probabiliste. Il devient ainsi possible de raisonner efficacement dans l'incertain, c'est-à-dire quand on possède des informations incomplètes sur un système qui n'est pas parfaitement déterministe. L'ordinateur accède par ces méthodes à des stratégies de décision très proches de celles employées par les humains.

Introduction

Dans les articles précédents de la rubrique Intelligence Artificielle (IA), nous avons abordé deux grands modes de raisonnement implémentables informatiquement. Nous nous sommes d'abord intéressés à l'IA des jeux de stratégie classiques [1], qui bénéficie d'une connaissance complète et parfaite : à tout moment, l'ordinateur connaît l'état du jeu (par exemple de l'échiquier) et peut prédire l'effet d'un coup (qu'il soit réalisé par lui-même ou par l'adversaire), même dans des jeux qui comportent une part de hasard comme le backgammon. Dans une telle situation, divers algorithmes sont envisageables et permettent d'explorer l'espace des possibles, c'est-à-dire l'arbre du jeu.

Nous avons ensuite étudié la situation moins favorable dans laquelle l'information n'est pas complète : dans les jeux de stratégie en temps réel (STR) par exemple, l'ordinateur n'a qu'une connaissance incomplète de l'état du jeu [2]. Il ne connaît pas l'emplacement exact des troupes du joueur, l'état de ses ressources, les mouvements qu'il vient d'amorcer, les constructions en cours, etc. Les solutions proposées dans [2] consistent à modéliser le comportement du joueur, essentiellement grâce à l'historique de son comportement. En repérant des comportements fréquents, par exemple, en détectant que le joueur favorise toujours la production de telle ou telle unité, l'ordinateur parvient à anticiper les actions du joueur, ce qui lui confère un avantage décisif.

Ces deux modes de raisonnement n'épuisent pas, loin s'en faut, toutes les possibilités ouvertes à l'ordinateur. Il manque en fait un des éléments principaux de l'intelligence humaine : la faculté de déduction. Quand deux joueurs humains s'affrontent dans un jeu STR, ils utilisent beaucoup cette faculté pour augmenter leur connaissance de l'état du jeu. Si on observe une unité d'un certain type, par exemple un chevalier, on peut en déduire de nombreuses informations sur la base de l'adversaire.

En effet, dans les jeux STR, les unités sont produites par des bâtiments spécialisés dont l'ordre de construction est déterminé par les règles du jeu. Les chevaliers peuvent nécessiter par exemple la construction d'une écurie qui n'est elle-même accessible qu'après la construction d'une forge ou d'un autre bâtiment. Le simple fait de voir un chevalier ennemi permet donc de déduire que le joueur adverse possède une écurie et disons une forge, ce qui renseigne de façon très importante sur l'état de sa base. Bien entendu, les mécanismes de déduction de ce type s'appliquent dans de très nombreux autres contextes, comme nous le verrons dans le présent article.

Il est fréquent, quand on parle de déduction, de penser aux systèmes experts. Ces programmes reproduisent le raisonnement formel humain, en enchaînant des règles logiques. Dans l'exemple considéré précédemment, un système expert devrait donc connaître les règles de production des unités et de construction des bâtiments, pour pouvoir déduire des faits observés (l'adversaire possède au moins une unité chevalier) l'état de la base (l'adversaire possède au moins une écurie et une forge). Les systèmes experts classiques sont cependant limités car ils travaillent avec du savoir exact, en se basant sur des règles logiques de la forme « si les conditions suivantes sont remplies, alors les faits suivants sont vrais ».

Or, en pratique, on cherche à faire beaucoup plus complexe. Dans le cas du jeu STR, on souhaite par exemple estimer les ressources de notre adversaire. Si on observe deux chevaliers, on peut déduire des choses certaines (la présence d'une écurie et d'une forge, ainsi qu'un nombre minimal de sources de nourriture), mais on peut aussi faire quelques hypothèses. L'expérience prouve en effet qu'il est rare que l'adversaire développe des chevaliers s'il n'a pas déjà un certain nombre de fantassins et d'archers, sauf si la partie n'a pas commencé depuis longtemps, auquel cas il est matériellement impossible pour l'adversaire d'avoir développé beaucoup d'unités. Le problème avec les systèmes experts est qu'il est difficile de donner un sens satisfaisant au mot « rare » ou à l'expression « un certain nombre » dans le cadre d'un raisonnement logique. Divers efforts ont été réalisés dans ce sens, essentiellement autour de la logique floue, ce qui conduit à une modification assez radicale des mécanismes du système expert.

Dans le présent article, nous nous intéressons à un type de modélisation pour lequel les incertitudes évoquées ci-dessus sont représentées à l'aide de loi de probabilité. Dans le cas du jeu STR, la présence chez l'adversaire de fantassins sachant qu'il dispose de chevaliers n'est plus certaine mais quantifiée par une grandeur comprise entre 0 et 1 qui mesure la chance d'occurrence de cet événement : en d'autres termes par une probabilité.

Les règles de logique déductive totalement déterministes dans le cas des systèmes experts sont conservées mais plongées dans un cadre probabiliste [3]. Ce type de modélisation assure une plus grande souplesse et une meilleure prise en compte de la complexité des phénomènes étudiés. Les modèles de ce type sont appelés des réseaux Bayésiens [4, 5].

Ils permettent de représenter la connaissance que l'on a d'un système (informatique, biologique, économique...) à travers, d'une part, des quantités aléatoires qui le composent, et, d'autre part, des liens d'influence entre ces différentes quantités. Un réseau Bayésien est ainsi schématisé par un graphe dans lequel les nœuds représentent des variables aléatoires et les liens des influences causales entre ces variables. Le graphe est acyclique : il ne contient pas de boucle.

Les domaines d'applications des réseaux Bayésiens et les types d'applications sont très variés. Une fois le modèle construit, on peut utiliser une partie de l'information disponible afin de prévoir, contrôler ou simuler le comportement du système. Une des raisons du succès actuel des réseaux Bayésiens, outre leur efficacité, est qu'ils peuvent être employés dans de très nombreux domaines comme l'industrie, la santé, la finance et même éventuellement le droit. Le système modélisé par un réseau Bayésien peut être le moteur d'une automobile, le patient d'une consultation médicale ou encore une fusée en plein vol. L'objectif est ici de montrer comment ces réseaux formalisent la notion de connaissance incertaine de façon mathématiquement valide tout en permettant le raisonnement dans ce cadre probabiliste. Dans une première partie, nous introduirons les concepts à travers l'étude d'un exemple complet. Puis, nous présenterons le cas général dans une deuxième partie.

Un exemple complet

Un problème d'alarme

Pour bien comprendre l'intérêt et les mécanismes des réseaux Bayésiens, nous allons commencer par étudier un exemple simple : nous considérons des alarmes installées dans des magasins et reliées à une société de surveillance. L'alarme se déclenche dans deux magasins simultanément (ou presque) et la société doit déterminer dans quel magasin envoyer une équipe en premier. La première difficulté dans ce problème est la conséquence des défauts des alarmes. En effet, il n'est pas rare que le passage d'un gros camion à côté d'un magasin déclenche intempestivement l'alarme. De plus, il arrive aussi que des cambrioleurs agiles parviennent à commettre leur larcin sans être détectés. De ce fait, quand l'alarme se déclenche, nous ne sommes pas certains qu'un vol soit bien en cours dans le magasin concerné. Cette réalité pourrait difficilement être prise en compte par un système expert dans lequel on aurait une règle de la forme « si alarme alors vol ». Au contraire, grâce à un réseau Bayésien, nous allons calculer la probabilité qu'un cambriolage soit en cours dans un magasin dont l'alarme vient de se déclencher, valeur sur laquelle la société de surveillance basera son plan d'action.

Modélisation graphique

Un réseau Bayésien est un instrument mathématique qui facilite le calcul des probabilités de différents évènements quand on possède une information partielle sur l'état du système considéré. Il s'agit plus précisément d'un modèle **graphique**, au sens où il s'appuie sur un graphe pour décrire le système. Dans notre exemple, le système est constitué des deux magasins. Pour chaque magasin, nous avons une alarme (en marche ou non) et des informations sur « l'environnement » du magasin : un gros camion vient-il de passer devant le magasin ? Des voleurs sont-ils en train de s'introduire dans le magasin ?

Ces différentes informations sont reliées selon un modèle causal qui ressemble aux règles logiques utilisées par un système expert : si un gros camion vient de passer devant le magasin, alors l'alarme risque de s'être déclenchée ; si des voleurs sont en train de s'introduire dans le magasin, alors l'alarme s'est sûrement déclenchée. Enfin, nous savons que certaines informations n'ont aucune raison d'être liées : il n'y a pas de raison pour que le passage d'un gros camion soit lié à la présence de voleurs, et vice et versa.

Cette structure constituée du modèle causal et des informations est résumée par un graphe, c'est-à-dire un ensemble de nœuds reliés par des arcs.

Chaque nœud correspond à une information, alors qu'un arc relie une information à une autre dans le sens supposé de causalité : si le graphe comporte un arc de A vers B, alors connaître la valeur de A donne des informations sur les valeurs possibles de B. La figure 1 donne le graphe correspondant à un magasin. Nous allons faire l'hypothèse que les magasins et les informations associées sont totalement indépendants et donc que le graphe global est la simple réunion de deux graphes associés à chacun des magasins, sans qu'on ajoute d'arc ou de nœud.

Modèle mathématique

Variable aléatoire

Pour utiliser le graphe de façon fructueuse et cohérente, il faut se placer dans un cadre mathématique précis (mais heureusement assez simple). L'idée fondamentale est d'adopter un langage probabiliste pour modéliser l'incertitude. Chaque information considérée dans le modèle est alors une **variable aléatoire**, c'est-à-dire une grandeur qui peut prendre plusieurs valeurs sans qu'on sache *a priori* quelle est sa valeur actuelle.

Dans l'exemple considéré, chaque variable est à valeurs booléennes : en effet, soit des voleurs sont présents (et donc Voleurs = vrai), soit aucun voleur n'est présent (et donc Voleurs = faux). Des généralisations sont possibles : on pourrait par exemple étudier le nombre de voleurs, ce qui donne comme valeur possible pour Voleurs n'importe quel entier positif.

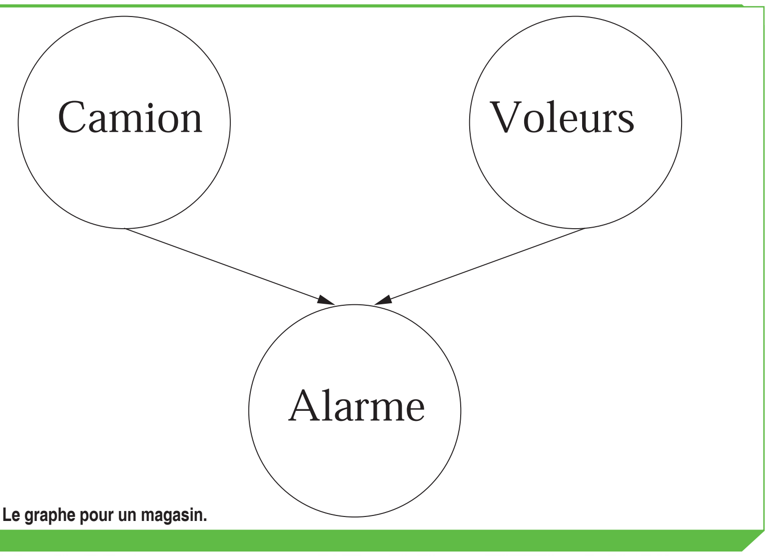


Fig. 1 : Le graphe pour un magasin.

Loi de probabilité

Une variable aléatoire est caractérisée par une **loi de probabilité**. Celle-ci représente notre seule connaissance sur la variable. En fait, la loi d'une variable aléatoire associe à chacune des valeurs possibles pour la variable une grandeur comprise entre 0 et 1 qui mesure la chance d'observer la variable dans l'état considéré : 0 est associé à une valeur impossible, alors que 1 correspond à une valeur certaine. Considérons par exemple la variable Camion. Si le magasin concerné est situé à proximité d'une route comportant un trafic important, il est fort probable que Camion = vrai, c'est-à-dire qu'un camion vienne de passer devant le magasin. La loi de Camion associe donc une valeur importante à la valeur vrai, par exemple 0.9, ce qui se note *P(Camion = vrai) = 0.9*.

Une condition importante vérifiée par les lois de probabilité est que la somme des probabilités associées à chacune des valeurs possibles pour une variable est égale à 1. Dans notre exemple, soit Camion = vrai, soit Camion = faux, on a donc :

$$P(Camion = vrai) + P(Camion = faux) = 1$$

Ainsi, *P(Camion = faux) = 0.1*. En pratique, on peut interpréter la probabilité associée à une valeur comme une **fréquence**. Dire *P(Camion = faux) = 0.1* signifie que si on regarde disons 100 fois devant le magasin pendant quelques instants, dans 10 cas aucun camion ne passera, alors que dans 90 un camion passera. On peut aussi considérer la probabilité comme une **croyance** *a priori* sur un évènement futur. Par exemple si nous considérons une pièce de monnaie quelconque, nous n'avons pas de raison de croire qu'elle est dissymétrique. Il est alors logique de supposer que *P(Pièce = pile) = P(Pièce = face) = 0.5*, au moins avant tout lancé. Il existe une vieille querelle de mathématiciens entre les tenant de la vision fréquentiste et ceux de la vision a prioriste (qu'on appelle aussi l'école Bayésienne, sans que cela n'ait de rapport direct avec les réseaux Bayésiens qui nous intéressent ici !). Nous ne nous engagerons pas sur ce terrain glissant et nous renvoyons les lecteurs intéressés par ce sujet à l'excellent et accessible ouvrage de Ian Hacking, *L'émergence de la probabilité* [6], qui présente l'histoire de cette branche des mathématiques et illustre les différents points de vue.

Quand on s'intéresse à plusieurs variables à la fois, on utilise la loi de probabilité **jointe**. Par exemple, *P(Camion = vrai, Voleurs = vrai)* désigne la probabilité d'avoir un passage de camion devant le magasin alors que celui-ci est en train d'être dévalisé par des voleurs. Bien qu'il existe des liens entre la loi jointe de plusieurs variables et les lois de chacune d'entre elles (qu'on appelle alors les lois **marginales**), il est en général impossible de construire la loi jointe à partir des lois marginales. Nous verrons justement que l'un des intérêt des réseaux Bayésiens est de rendre les calculs possibles.

Conditionnement

Dans certaines situations, il est très difficile d'attribuer une loi de probabilité à une variable sans informations supplémentaires. C'est le cas dans notre exemple pour la variable Alarme. Comment en effet attribuer une valeur à *P(Alarme = vrai)* ? L'alarme est un dispositif presque déterministe et on peut raisonnablement penser que la connaissance des conditions extérieures (présence de voleurs et/ou passage d'un camion) suffisent pour obtenir la valeur de la variable. Pourquoi alors parler d'aléa et de probabilités ? En fait, ces dernières interviennent à deux moments. Tout d'abord, l'état de l'alarme n'est pas parfaitement connu même si on sait que des voleurs sont présents dans le magasin par exemple : en effet, comme nous l'avons rappelé plus tôt, il arrive que des voleurs ne soient pas détectés. De plus, comme notre connaissance des conditions extérieures (les variables Camion et Voleurs) n'est que probabiliste, même pour une alarme parfaitement déterministe, la valeur de Alarme reste fondamentalement aléatoire !

Pour formaliser ces constatations, nous utilisons la notion de **probabilités conditionnelles**. Le principe est très simple : il est parfois plus logique de spécifier la loi de probabilité suivie par une variable aléatoire **sachant les valeurs de certaines autres variables**. En général, ceci correspond à un lien de causalité probabiliste entre la variable étudiée et les variables dont on fixe les valeurs. Considérons le cas de l'alarme. Si celle-ci est très performante, elle détecte la plupart des voleurs.

On peut donc supposer que la probabilité de Alarme = vrai est proche de 1 quand Voleurs = vrai. Cependant, une alarme sensible se déclenche aussi facilement quand un camion passe à proximité, donc la probabilité de Alarme = vrai est non nulle quand Camion = vrai. Il semble logique de dire que les effets sont ici cumulatifs et donc que l'alarme se déclenche très facilement quand des voleurs sont présents et qu'un camion passe devant de magasin !

Dans le même ordre d'idée, on peut supposer que l'alarme ne se déclenche pas s'il n'y a ni voleur, ni camion. Pour donner une probabilité conditionnelle, on utilise la notation suivante :

$$P(Alarme = vrai | Camion = Vrai, Voleurs = vrai) = 1$$

La variable considérée est à gauche de la barre verticale, alors que les variables dont les valeurs sont fixées sont situées à droite (ce sont les variables par rapport auxquelles on conditionne la première). La notation se lit de la façon suivante : « la probabilité d'avoir Alarme = vrai sachant que Camion = vrai et que Voleurs = vrai est de 1. » Une loi de probabilité conditionnelle respecte la règle de somme à 1. Plus précisément, si on fixe les valeurs de certaines variables, la somme des probabilités conditionnelles des valeurs possibles pour la variable étudiée doit faire 1. Par exemple, comme *P(Alarme = vrai | Camion = vrai, Voleurs = vrai) = 1*, alors *P(Alarme = faux | Camion = vrai, Voleurs = vrai) = 0*, car Alarme ne peut prendre que deux valeurs.

Le tableau 1 résume en termes de probabilités conditionnelles la discussion sur l'alarme proposée précédemment. Il indique *P(Alarme = vrai | ...)* pour toutes les combinaisons possibles des valeurs des deux autres variables considérées. La valeur de *P(Alarme = faux | ...)* s'en déduit par la règle de somme à 1.

	Camion = vrai	Camion = faux
Voleurs = vrai	1	0.98
Voleurs = faux	0.05	0

Tableau 1 : Probabilité de Alarme = vrai sachant les valeurs des autres variables.

Une alarme moins sensible détecte moins souvent les voleurs, mais en contrepartie ne se déclenche pas intempestivement lors du passage d'un camion.

Mathématiquement, ceci se traduit par exemple par les probabilités conditionnelles données par le tableau 2.

	Camion = vrai	Camion = faux
Voleurs = vrai	0.96	0.94
Voleurs = faux	0.02	0

Tableau 2 : Probabilité de Alarme = vrai sachant les valeurs des autres variables.

Le réseau Bayésien

Pour obtenir un réseau Bayésien qui modélise un problème, il faut donc déterminer les variables aléatoires concernées, puis établir des liens de causalité incertains entre celles-ci, qu'on résume par un graphe. La troisième étape consiste à spécifier pour chaque variable sa loi de probabilité conditionnelle. Le conditionnement de la variable A se fait obligatoirement par rapport aux variables qui sont reliées à A par un arc dans le sens variable vers A (il s'agit donc des prédécesseurs du nœud A dans le graphe, aussi appelés **parents** de A). Ceci correspond à la notion de causalité retenue dans le modèle probabiliste.

Dans notre exemple, il s'agit donc de fixer les lois suivantes :

- 1. comme aucune variable n'est reliée à Camion, la loi à fournir pour cette variable n'est pas conditionnelle, il s'agit donc simplement de *P(Camion = vrai)*, soit par exemple *P(Camion = vrai) = 0.9* dans un lieu très fréquenté et *P(Camion = vrai) = 0.05* dans un lieu plutôt désert ;
- 2. de la même façon, Voleurs ne possède pas de parent dans le graphe et on doit donc donner *P(Voleurs = vrai)*. Dans une zone de forte criminalité, on peut avoir *P(Voleurs = vrai)=0.05* et par exemple *P(Voleurs = vrai)=0.01* pour une zone plus sûre ;
- 3. enfin, comme Alarme possède deux prédécesseurs, Camion et Voleurs, nous devons préciser la loi de Alarme conditionnellement à Camion et Voleurs, ce que nous avons fait pour deux exemples dans les tableaux 1 et 2.

Calculer avec le réseau

Probabilités non conditionnelles

Nous avons déjà indiqué qu'il est assez difficile de déterminer la loi de certaines variables aléatoires sans introduire de conditionnement. C'est le cas par exemple de la variable Alarme. Que dire en effet de *P(Alarme = vrai)* ? Une des utilités des réseaux Bayésiens est qu'ils permettent de calculer ce genre d'information à partir des probabilités conditionnelles qui décrivent le réseau. Pour ce faire, on utilise trois propriétés fondamentales des probabilités :

- 1. la propriété des **lois marginales** : quand on cherche à calculer P(Alarme = vrai), une première étape consiste à exprimer cette probabilité comme une somme de probabilités portant sur des groupes de variables : on obtient la loi marginale (la loi d'une variable) à partir de la loi jointe. Considérons d'abord le cas abstrait de deux variables aléatoires A et B à valeurs booléennes. On montre que l'expression suivante est vraie :

$$P(A=vrai)=P(A=vrai, B=vrai)+P(A=vrai, B=faux)$$

Plus généralement, si B peut prendre plusieurs valeurs, on obtient *P(A=vrai)* en réalisant la somme des *P(A=vrai, B=b)* pour toutes les valeurs possibles de b. Ici, on obtient donc *P(Alarme = vrai)* en considérant la somme de toutes les possibilités de la forme *P(Alarme = vrai, Camion = c, Voleurs = v)*. Pour simplifier la suite du texte, nous noterons C = c pour Camion = c et V = v pour Voleurs = v.

- 2. la définition des **probabilités conditionnelles** : nous avons pour l'instant considéré que les probabilités conditionnelles étaient des données du problème. En fait, dans certaines situations, elles sont calculées à partir de leur définition qui est la suivante (pour deux variables A et B) :

$$P(A=a|B=b)=\frac{P(A=a, B=b)}{P(B=b)}$$

Dans notre cas, celle-ci signifie qu'on a :

$$P(Alarme=vrai, C=c, V=v)=P(Alarme=vrai|C=c, V=v)P(C=c, V=v)$$

Nous avons donc fait un pas décisif vers la solution car nous connaissons les *P(Alarme = vrai | C = c, V = v)*, par hypothèse. Reste donc à calculer les *P(C = c, V = v)*.

- 3. la notion d'**indépendance** : on dit que deux variables aléatoires sont indépendantes quand connaître la valeur de l'une ne donne aucune information sur l'autre, ce qui se traduit par la formule suivante :

$$P(A=a, B=b)=P(A=a)P(B=b)$$

Quand *P(B = b)* n'est pas nul, ceci se traduit aussi sous la forme suivante :

$$P(A=a|B=b)=P(A=a)$$

Certaines règles des réseaux Bayésiens permettent de déterminer quelles sont les variables indépendantes en fonction du graphe et des informations connues. Dans notre application, Camion et Voleurs sont indépendants. En d'autres termes, si je sais que des voleurs sont présents dans un magasin, cela ne me donne aucune information sur le passage récent d'un camion devant celui-ci (et vice versa), comme nous l'avons déjà dit en introduction. Cela signifie donc que *P(C = c, V = v) = P(C = c)P(V =v)*. Nous avons donc ainsi toutes les informations nécessaires au calcul de *P(Alarme = vrai)*.

Considérons à titre d'illustration le cas d'une zone relativement sûre, i.e. P(V = vrai) = 0.005, et peu fréquentée par les camions, soit P(C = vrai) = 0.02. Le calcul des P(C = c, V = v) se récapitule dans le tableau 3.

	Camion = vrai	Camion = faux
Voleurs = vrai	1,00E-4	4,90E-3
Voleurs = faux	1,99E-2	9,751E-1

Tableau 3 : Probabilité de Camion = c, Voleurs = v

Pour obtenir par exemple *P(Alarme = vrai, C = vrai, V = faux)*, il suffit de multiplier *P(Alarme = vrai | C = vrai, V = faux) = 0.02* (ce qui correspond à l'alarme décrite dans le tableau 2) par *P(C = vrai, V = faux) = 0.0199*, ce qui nous donne *0.000398*. En réalisant la somme des probabilités correspondant aux quatre cas possibles, on obtient *P(Alarme = vrai) = 0.0051*. Si on avait fait l'hypothèse simpliste que l'alarme détecte toujours les voleurs et n'est jamais déclenchée par un camion, on aurait obtenu une probabilité *P(Alarme = vrai) = 0.005*, ce qui peut sembler très proche.

Cependant, ces valeurs presque identiques masquent le fait que dans le cas du calcul complet, l'alarme peut être déclenchée par un camion, ce qui change tout pour l'entreprise de télé-surveillance.

Inférence

Le plus intéressant dans les réseaux Bayésiens n'est pas de pouvoir calculer la loi de probabilité de toutes les variables qui le constituent. Comme nous l'avons vu dans l'exemple numérique traité précédemment, les résultats ainsi obtenus sont utiles pour certaines applications (par exemple déterminer une police d'assurance en fonction du risque de sinistre dans une zone) mais ne sont d'aucune aide dans d'autres : si je sais que l'alarme vient de se déclencher dans un magasin, l'information *P(Alarme = vrai) = 0.0051* m'est totalement inutile pour estimer le risque que l'alarme ait été déclenchée par des voleurs.

En fait, dans le problème qui nous intéresse ici est le calcul de *P(Voleurs = vrai | Alarme = vrai)*, ainsi d'ailleurs que *P(Camion = vrai | Alarme = vrai)*. En effet, pour prendre une décision, la société de télé-surveillance doit connaître le risque qu'elle prend en se rendant d'abord dans le premier ou le second magasin. De plus, la valeur de la variable Alarme est connue et peut donc être utilisée pour obtenir de l'information.

Diverses stratégies de calcul sont envisageables pour obtenir les valeurs recherchées, toujours en se basant sur les trois propriétés introduites précédemment. On remarque cependant qu'on cherche ici à renverser la causalité. En effet, on connaît naturellement la loi de Alarme sachant Voleurs et Camion, et on cherche ici à obtenir au contraire la loi de Voleurs sachant Alarme. Dans une telle situation, la **règle de Bayes**, dérivée de la définition des probabilités conditionnelles, est très utile. Si A et B sont deux variables aléatoires, la règle s'écrit :

$$P(A=a|B=b)=\frac{P(B=b|A=a)P(A=a)}{P(B=b)}$$

En notant A comme résumé de Alarme = vrai, on obtient ainsi :

$$P(Voleurs=vrai|Alarme=vrai)=\frac{P(A|V=vrai)P(V=vrai)}{P(A)}$$

Nous venons de calculer P(A) et nous connaissons P(V = vrai) par hypothèse. Il nous reste donc à calculer *P(Alarme = vrai | Voleurs = vrai)*. Pour ce faire, on utilise la règle sur les loi marginales, appliquées à la loi de Alarme sachant Voleurs = vrai, ce qui nous donne :

$$P(A|V=vrai)=P(A,C=vrai|V=vrai)+P(A,C=faux|V=vrai)$$

On applique ensuite la définition de probabilités conditionnelles à chaque terme. On a par exemple :

$$P(A,C=vrai|V=vrai)=\frac{P(A,C=vrai,V=vrai)}{P(V=vrai)}$$

ce qui se transforme en :

$$P(A,C=vrai|V=vrai)=\frac{P(A|C=vrai,V=vrai)P(C=vrai,V=vrai)}{P(V=vrai)}$$

et enfin en :

$$P(A,C=vrai|V=vrai)=P(A|C=vrai,V=vrai)P(C=vrai)$$

car Voleurs et Camion sont des variables indépendantes. En conduisant les calculs jusqu'au bout, on obtient *P(Alarme = vrai | Voleurs = vrai)=0.9404*. En combinant ce résultat avec ceux de la section précédente, on obtient *P(Voleurs = vrai | Alarme = vrai)= 0.9220*.

Des calculs similaires conduisent à la valeur *P(Camion = vrai | Alarme = vrai)= 0.0969*. Il est donc beaucoup plus probable que l'alarme ait été provoquée par des voleurs que par un camion.

Considérons maintenant le cas d'un magasin équipé d'une alarme très sensible (tableau 1) dans une zone de faible criminalité (*P(V = vrai) = 0.005*), mais relativement passante (*P(C = vrai) = 0.05*).

En reproduisant les calculs exposés pour le cas précédent avec les nouvelles valeurs numériques, on obtient *P(Voleurs = vrai | Alarme = vrai)=0.6635* et *P(Camion = vrai | Alarme = vrai)= 0.3703*.

Bien entendu, on reste dans une situation où il est plus probable que l'alarme ait été déclenchée par des voleurs, mais son déclenchement par un camion n'est plus aussi peu probable qu'avant.

Si on compare les deux magasins, il semble plus logique d'envoyer d'abord une équipe d'intervention dans le premier, car il est à la fois plus probable que des voleurs soient en train de le dévaliser et moins probable que l'alarme ait été déclenchée par un camion.

Le cas général

Un réseau Bayésien

De manière générale, un réseau Bayésien est défini par les trois éléments suivants :

- 1. Un ensemble de *n* variables aléatoires A_1, A_2, ..., A_n ;
- 2. Un ensemble de relations causales entre ces différentes variables :
  - pour chaque variable A_i, on dispose de l'ensemble de ses parents, ensemble noté $pa(A_i)$. Il s'agit du sous-ensemble des variables qui causent **directement** A_i, c'est-à-dire celles qui ont une influence causale sur A_i sans intermédiaire(s) (cette notion sera définie plus précisément dans la suite du texte) ;
- 3. Les lois de probabilité des variables sachant leurs parents :
  - pour simplifier les notations, nous désignons par $P(i)_1$, $P(i)_2$, ..., $P(i)_{n(i)}$ les parents de A_i (alors qu'il s'agit en fait de certains A_k pour des *k* bien choisis). Dans cette liste, *n(i)* est le nombre d'éléments de $pa(A_i)$. Pour chaque variable A_i, on dispose de la loi conditionnelle :

$$P(A_i=a_i|P(i)_1=p_1,\ldots,P(i)_{n(i)}=p_{n(i)}).$$

Dans notre exemple introductif, on retrouve les variables aléatoires Camion, Voleurs et Alarme, associées dans le graphe de la figure 1 et les probabilités spécifiées dans la description, en particulier celles qui décrivent l'alarme (cf les tableaux 1 et 2). Notons que la manière la plus simple de représenter un réseau Bayésien est de dessiner le graphe orienté associé à la notion d'influence.

La propriété fondamentale des réseaux Bayésiens ainsi définis réside dans le fait que la loi de probabilité jointe des variables aléatoires A_1, A_2, ..., A_n du réseau est égale au produit des lois conditionnelles de toutes les variables sachant leurs parents, ce qui s'écrit :

$$P(A_1=a_1,\ldots,A_n=a_n)=\prod_{i=1}^{n} P(A_i=a_i|P(i)_1=p_1,\ldots,P(i)_{n(i)}=p_{n(i)}).$$

Cette factorisation particulière a naturellement des conséquences en termes de dépendance et d'indépendance (conditionnelle ou pas) entre les variables du réseau. De plus, elle permet de définir un algorithme (certes basique) de calcul automatique des diverses lois de probabilités dont on peut avoir besoin pour exploiter le réseau.

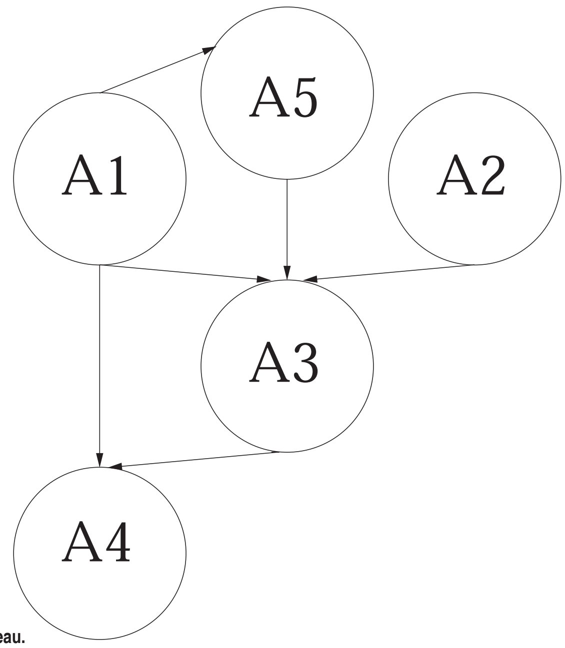


Fig. 2 : Un exemple de réseau.

Avant de nous plonger dans les détails pratiques, étudions un exemple simple. Considérons un réseau formé de 5 variables dont les relations causales sont précisées par le graphe de la figure 2. La formule générale ci-dessus se traduit alors en la version suivante :

$$P(A_1=a_1,\ldots,A_5=a_5)=P(A_1=a_1)\,P(A_2=a_2)\,P(A_3=a_3|A_1=a_1,A_2=a_2,A_5=a_5)\,P(A_4=a_4|A_1=a_1,A_3=a_3)\,P(A_5=a_5|A_1=a_1)$$

On constate que cette formule ne fait apparaître que des lois de probabilité qui doivent être spécifiées dans la description du réseau (comme indiqué au point 2). En appliquant la formule générale à l'exemple de l'alarme, nous obtenons la loi jointe suivante :

$$P(Alarme=a,Camion=c,Voleurs=v)=P(Alarme=a|Camion=c,Voleurs=v)\,P(Camion=c)\,P(Voleurs=v)$$

Cette formule est exactement celle que nous avons utilisée dans l'exemple en la retrouvant par un raisonnement probabiliste.

Inférence sur un réseau

Avant de nous attaquer au problème complexe de la construction d'un réseau Bayésien, supposons celui-ci donné. Nous avons vu dans l'exemple introductif qu'on peut utiliser le réseau pour calculer des lois de probabilité pour les variables qui nous intéressent, sachant les valeurs prises par d'autres variables. Le problème est que nous avons utilisé un raisonnement probabiliste difficilement implémentable de manière systématique.

Cependant, en théorie, le calcul général se programme sans difficulté majeure en appliquant la formule de factorisation de la loi jointe, la propriété des lois marginales et la définition des probabilités conditionnelles. Reprenons à titre d'exemple le calcul de la loi de probabilité de la variable Voleurs sachant que Alarme = vrai.

D'après la définition des probabilités conditionnelles, nous avons :

$$P(Voleurs=v|Alarme=vrai)=\frac{P(Voleurs=v,Alarme=vrai)}{P(Alarme=vrai)}$$

D'après la propriété des lois marginales, *P(Voleurs = v, Alarme = vrai)* s'obtient par simple somme des *P(Voleurs = v, Alarme = vrai, Camion = c)* pour toutes les valeurs possibles de *c* (ici vrai et faux). De même *P(Alarme = vrai)* s'obtient par somme des *P(Voleurs = v, Alarme = vrai, Camion = c)* pour toutes les valeurs de *v* et de *c*.

Enfin, comme nous venons de le voir, la formule de factorisation permet de calculer *P(Voleurs = v, Alarme = a, Camion = c)* pour n'importe quelles valeurs de *v*, *a* et *c*, à partir des informations fournies lors de la construction du réseau.

Il est clair que le raisonnement que nous venons de produire s'applique dans le cas général. En effet, nous cherchons toujours une loi conditionnelle et donc une probabilité de la forme *P(A = a, B = b, ... | X = x, Y = y, ...)*.

Une telle grandeur s'obtient par division de deux probabilités non conditionnelles qu'on obtient par simple somme à partir de la loi jointe globale *P(A = a, B = b, ... , X = x, Y = y, ...)* qui est elle-même donnée par la formule de factorisation.

Le gros défaut de l'algorithme simple ainsi obtenu est son coût. En effet, dans le cas le pire, on peut être amené à sommer un nombre considérable de termes de la loi jointe.

En fait, la loi jointe grossit exponentiellement avec le nombre de variables : pour la connaître complètement, il faut en effet calculer les *P(A = a, B = b, C = c, ...)* pour toutes les valeurs possibles de *a*, *b*, *c*, ...

Le coût vient du fait que toutes les combinaisons doivent être étudiées. Si on a *n* variables binaires, on obtient ainsi 2^n combinaisons. Pour *n = 10*, cela ne pose pas de problème, mais pour *n = 30*, les choses se compliquent radicalement.

De plus, il est fréquent d'utiliser des variables avec plus de deux valeurs possibles, ce qui ne fait qu'augmenter les problèmes.

Fort heureusement, il existe des algorithmes dits de propagation qui exploitent la structure du réseau et permettent de réduire considérablement le temps de calcul. Le plus connu est le *Junction Tree Algorithm* qui est un cas particulier de *Belief Propagation*.

La *Belief Propagation* ou propagation des croyances est une méthode générale permettant de résoudre des problèmes d'inférence dans des domaines aussi variés que la physique statistique, l'intelligence artificielle, le décodage de codes correcteurs d'erreurs, la vision par ordinateur, etc.

Elle a été découverte séparément dans ces différents domaines sous les noms suivants : Algorithme de *Viterbi*, *Filtre de Kalman*, Algorithmes de *décodage itératif* pour les codes de *Gallager, Turbocodes,* etc.

Elle s'applique dès que la structure probabiliste considérée peut être représentée par un graphe. L'état le plus probable des nœuds du graphe est alors recherché.

Ce type d'algorithme consiste à calculer itérativement les lois de probabilités des états des nœuds du graphe à partir des messages locaux envoyés d'un nœud à ses voisins. Les lecteurs intéressés par les algorithmes d'inférence pourront se reporter à [4, 9].

Diverses implémentations (libres ou non) du *junction tree* ou d'autres algorithmes efficaces sont listées dans [10]. Nous avons réalisé nos calculs avec le p*ackage Grappa* [11] conçu pour le logiciel libre de statistique *R* [12].

Notons pour conclure que dans certaines situations, aucune simplification des calculs n'est possible et que le réseau Bayésien correspondant est donc difficile à exploiter en un temps de calcul raisonnable.

Retour sur les alarmes

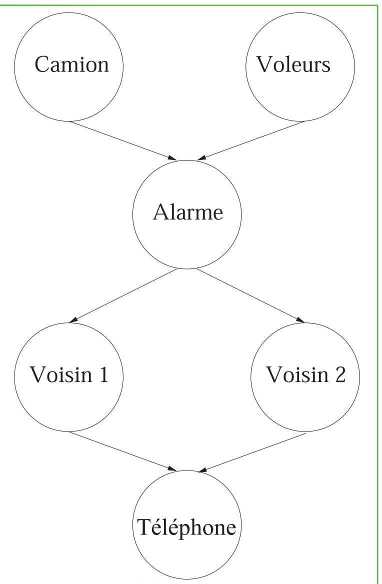


Fig. 3 : Voisins et téléphone.

Pour illustrer les possibilités des réseaux Bayésiens généraux, nous allons étendre l'exemple introductif en considérant un seul magasin dont l'alarme n'est plus reliée à une société de surveillance mais à une sirène qui émet un signal sonore dans la rue. Supposons par ailleurs que seulement deux voisins proches du magasin peuvent entendre l'alarme. Si l'un des deux voisins entend l'alarme, il y a une forte probabilité pour qu'il téléphone au gérant du magasin. Le réseau Bayésien correspondant est représenté par la figure 3.

Pour être tout à fait précis, le réseau retenu introduit une complication supplémentaire, représentée par la variable Téléphone. L'idée est que le gérant du magasin ne reçoit pas directement l'appel d'un voisin : un de ses collaborateurs prend l'appel et dans l'urgence ne pense pas à noter le nom du voisin.

De ce fait, la seule information connue par le gérant est qu'un des deux voisins a entendu la sirène du magasin. Il n'est donc pas possible de prendre directement en compte les caractéristiques des voisins, comme par exemple le fait que l'un d'entre eux soit paranoïaque et ait appelé dans le passé pour rien, ou au contraire le fait qu'un des voisins soit sourd et n'ait donc pas appelé alors que l'alarme s'était déclenchée.

Une fois prévenu de l'appel, le gérant doit prendre une décision : appeler la police ou se rendre seul au magasin. D'un côté, s'il fait se déplacer les policiers pour rien, ces derniers risquent éventuellement de ne plus se rendre rapidement à son magasin lors d'un éventuel prochain appel de détresse. D'un autre côté, si des voleurs sont présents dans le magasin, il est dangereux pour le gérant de s'y rendre seul. Ainsi, pour prendre la décision d'appeler ou non la police, il lui est fort utile d'évaluer la probabilité que des voleurs se trouvent dans son magasin sachant que l'un ou l'autre des deux voisins lui a téléphoné.

Pour spécifier complètement le réseau, il faut préciser les lois conditionnelles *P(Voisin 1 = v | Alarme = a)* et *P(Voisin 2 = v | Alarme = a)* : c'est ici qu'on modélise la paranoïa ou la surdité des voisins !

On doit aussi donner la loi de Téléphone sachant les valeurs des deux voisins, mais dans un premier temps, on peut se contenter d'un modèle déterministe précisant que dès qu'un voisin appelle, la variable téléphone vaut vrai. On peut aussi compliquer le modèle en supposant que le collaborateur du gérant du magasin note le nombre d'appels reçus, la variable Téléphone prenant ainsi les valeurs 0, 1 ou 2 selon les valeurs de Voisin 1 et Voisin 2.

On peut enfin modéliser l'oubli de l'appel par le collaborateur en donnant une valeur non nulle pour *P(Téléphone = 0 | Voisin 1 = vrai, Voisin 2 = vrai).*

Les deux probabilités conditionnelles intéressantes sont bien entendu *P(Voleurs = vrai | Téléphone = vrai)* et *P(Camion = vrai | Téléphone = vrai).* Comme nous l'avons indiqué dans la section précédente, les calculs se font « simplement » à partir des différentes lois conditionnelles en utilisant les différentes propriétés des lois de probabilité.

Construction d'un réseau Bayésien

Choix des variables

Lorsque l'on souhaite modéliser un système à l'aide d'un réseau Bayésien, il faut, dans un premier temps, déterminer l'ensemble des variables pertinentes. Ces variables peuvent être classées en trois catégories :

- **1. variables d'intérêt** : ce sont les variables que l'on n'observe pas et dont on souhaite prédire le comportement. Dans l'exemple de l'alarme, il s'agit des variables Camion et Voleurs ;

- **2. variables d'information** : ce sont les variables qui sont observées et qui apportent de l'information sur les variables d'intérêt. Dans l'exemple de l'alarme, il s'agit de la variable Alarme. Quand on passe à l'exemple plus complexe, Alarme devient une variable de modélisation et la nouvelle variable Téléphone est la seule variable d'information ;

- **3. variables de modélisation** : ce sont les variables introduites dans le modèle dans le but soit de mieux spécifier les probabilités conditionnelles, soit de mieux représenter les relations d'indépendance entre les variables d'information et les variables d'intérêt. Elles doivent être introduites dans le modèle avec grande précaution. En effet, si elles sont présentes uniquement dans le souci de raffiner ce dernier, elles peuvent considérablement en dégrader les performances. Dans l'exemple étendu, Alarme, Voisin 1 et Voisin 2 sont des variables de modélisation. Elles sont ici indispensables car il n'est pas possible de spécifier le réseau en donnant directement la loi de Téléphone conditionnellement à Voleurs et Camion, tant une telle loi serait arbitraire.

La notion de causalité

La deuxième tâche importante dans la spécification d'un réseau Bayésien consiste à faire apparaître des liens de cause à effet entre les variables qui constituent le système auquel on s'intéresse, ce qui n'est pas toujours facile. Pour fixer le vocabulaire, nous dirons dans un réseau Bayésien que A_j cause **directement** A_i si la variable A_j est reliée par une flèche à la variable A_i, dans le sens A_j vers A_i : dans l'exemple des magasins, Voleurs cause donc directement Alarme. On dira plus généralement que A_j cause A_i s'il existe un chemin (une suite de nœuds dans le graphe) qui mène de A_j vers A_i. Par exemple, Camion cause Téléphone dans l'exemple étendu.

Un point important à bien comprendre est que la notion de causalité directe ou non n'est pas uniquement liée à la notion d'influence. En effet, la variable A_j a une influence sur la variable A_i, si connaître la valeur de A_j donne des informations sur les valeurs possibles de A_i et donc sur sa loi de probabilité. Techniquement, cela signifie que la loi de A_i est différente de la loi conditionnelle de A_i sachant A_j, c'est-à-dire exactement le contraire de la définition d'indépendance entre deux variables donnée précédemment. De ce fait, si A_j a une influence sur A_i alors A_i a une influence sur A_j : en effet, la définition statistique de l'indépendance donne des rôles symétriques aux variables.

Ce n'est bien entendu pas le cas en termes de causalité. Ainsi, pour que A_j cause directement A_i il faut, d'une part, que A_j ait une influence sur la variable A_i et, d'autre part, que la manière la plus naturelle de spécifier cette influence s'effectue par l'intermédiaire de la loi conditionnelle de A_i sachant A_j.

Dans de nombreux cas, cela est assez naturel. Par exemple, dans l'exemple de l'alarme, l'influence des variables Camion et Voleurs sur la variable Alarme s'écrit naturellement par la loi de probabilité conditionnelle de Alarme sachant Camion et Voleurs, et non l'inverse. Cependant, comme nous l'avons vu dans les calculs, il est parfaitement possible de calculer la probabilité de *Voleurs = vrai* sachant *Alarme = vrai,* c'est le but ! Malheureusement, il existe des cas où il n'est pas possible de diriger ainsi le sens de l'influence, on ne peut pas utiliser les réseaux Bayésiens. Pour pallier cette difficulté, des modèles prenant en compte des influences causales et non causales ont été développés, il s'agit des Chain Graph Models [7] qui ne seront pas abordés ici.

Par ailleurs et dans le même ordre d'idées, les réseaux Bayésiens ne concernent que les systèmes pour lesquels les liens de causalité sont univoques et ainsi clairement définis. Aucun cycle de causalité ne doit apparaître. Si A_j cause A_i (directement ou non), il est impossible que A_i cause A_j (directement ou non). Ainsi, le graphe associé à un réseau Bayésien est un graphe acyclique dirigé (*Directed Acyclic Graph* ou *DAG* en anglais), dans lequel il est impossible d'emprunter un chemin en suivant les flèches, qui partirait de A_j pour y revenir. En effet, si cela était le cas et si la variable A_i se trouvait sur le chemin, alors on pourrait dire que A_j cause A_i et inversement.

Lois conditionnelles

Une fois la structure causale choisie, il faut spécifier les lois de probabilités de toutes les variables du modèle conditionnellement à leur(s) parent(s). On les appelle les lois *a priori* du modèle.

Elles reflètent soit les croyances d'experts, soit les caractéristiques du système. Dans l'exemple de l'alarme, la loi de probabilité de la variable Alarme conditionnellement aux variables Camion et Voleurs est fournie par la société de surveillance et caractérise la nature de l'alarme.

2 SITES INCONTOURNABLES

Toute l'actualité du magazine sur :

www.gnulinuxmag.com

Abonnements et anciens numéros en vente sur :

www.ed-diamond.com

Pour établir une telle loi, on peut recourir à l'expérimentation. L'idée est d'observer le système pendant un certain temps, en notant l'état de toutes les variables du réseau à chaque instant d'observation.

En interprétant les probabilités d'une façon fréquentiste, on peut alors donner une estimation des lois par simple comptage.

Dans le cas de l'alarme, on imagine facilement un dispositif expérimental simple : on installe une alarme dans un magasin « type », puis on fait passer plusieurs fois un camion devant le magasin.

On note combien de passages provoquent un déclenchement de l'alarme. On estime ainsi *P(Alarme = vrai | Camion = vrai, Voleurs = faux)* comme le rapport entre le nombre de passages ayant déclenché l'alarme et le nombre total de passage. Si on a étudié 50 passages avec 3 déclenchements, on estime donc la probabilité à 0.06.

Les variables des mots n'ont aucun lien entre elles (c'est ici que le modèle est simpliste) et sont toutes causées directement par la variable Spam. Le figure 4 présente le graphe caractérisant un tel réseau.

Pour spécifier complètement le réseau, il faut donner la loi de probabilité de la variable Spam. Celle-ci peut être évaluée par simple comptage dans la boîte mail considérée. Puis, il faut donner la probabilité conditionnelle *P(Mot = vrai | Spam = vrai)* pour chaque mot envisageable. Encore une fois, celle-ci peut être évaluée par simple comptage.

On se donne une boîte *mail* dans laquelle chaque courrier indésirable est marqué spam. A partir de cette boîte, on détermine un dictionnaire des mots les plus fréquents (sans tenir compte pour l'instant de la nature des messages). On calcule ensuite pour chaque mot le nombre de courriers indésirables et normaux dans lequel il apparaît, soit respectivement *s* et *h* ces nombres.

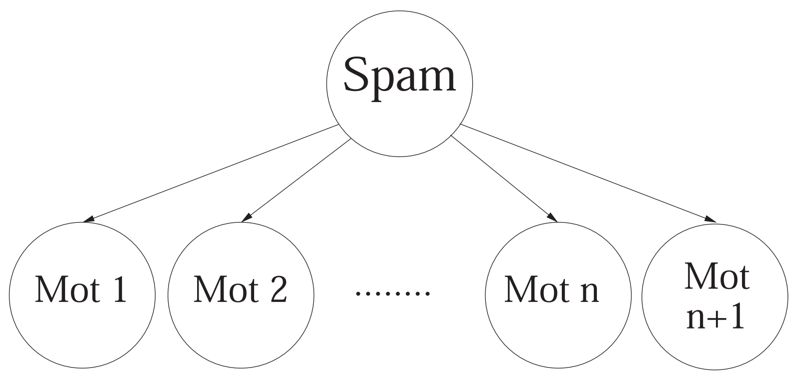


Fig. 4 : Graphe pour le filtrage de spam.

Ce type de construction est utilisé dans les méthodes de filtrage de spam par approche bayésienne (cf [12]). Le réseau Bayésien utilisé pour les filtrages de spam classiques est très simple (voire simpliste) : on considère une variable Spam qui indique la nature du *mail* (c'est la variable d'intérêt).

Cette variable est parent d'un très grand nombre de variables, en fait une variable par mot envisageable dans un *email*.

Chaque variable correspondant à un mot peut prendre la valeur vrai ou faux, selon que le mot est ou non présent dans l'email considéré (il s'agit donc d'une variable d'information).

Si *S* désigne le nombre total de courriers indésirables et *H* le nombre total de courriers normaux, *P(Mot = vrai | Spam = vrai)* peut être évaluée par *s/S* et *P(Mot = vrai | Spam = faux)* par *h/H*.

Une fois le modèle complètement décrit, il est facile de calculer *P(Spam = vrai | contenu du mail)* en appliquant les règles de calcul exposées plus haut. On peut ensuite rejeter un mail dont la probabilité d'être un spam est élevée. Malgré la simplicité du réseau Bayésien utilisé, cette technique fonctionne extrêmement bien en pratique. Elle a d'ailleurs été appliquée dans d'autres contextes comme la reconnaissance de formes en général sous le nom de **classification bayésienne naïve**.

Construction automatique

Une fois qu'on a spécifié la liste des variables entrant dans le modèle, on dispose de procédures automatiques qui permettent de choisir la structure de relation causale la plus pertinente.

Encore une fois, on a recours à l'expérience. Supposons que l'on ait observé le système pendant un certain temps et noté l'état de toutes les variables observables.

On peut alors fournir une estimation par simple comptage de la loi de probabilité jointe de ces variables. C'est ainsi que la technique de choix automatique de structure causale consiste à :

- 1. calculer, pour tous les réseaux possibles, la loi de probabilité jointe des variables observables ; pour un réseau donné, cette loi est déterminée par simple sommation de la loi de probabilité jointe de toutes les variables (loi donnée par la structure du réseau) sur les variables non observées ;
- 2. choisir le réseau pour lequel la loi de probabilité jointe des variables observables précédemment calculée est la plus proche de celle estimée par l'expérience.

Il s'agit de la technique dite du *Batch Learning*. Cette méthode est très bien implémentée dans le *package Deal* [14] du logiciel libre de statistique *R*.

Dans la majorité des cas, il est impossible de calculer la loi de probabilité jointe pour tous les réseaux possibles.

On utilise alors diverses heuristiques de simplification, comme par exemple l'algorithme d'élimination descendante de liens. Cette méthode consiste à partir du réseau contenant toutes les relations causales possibles et à éliminer progressivement les liens les moins pertinents.

Un autre point important est que dans la procédure de choix systématique, il est indispensable de définir une mesure de distance entre deux lois de probabilité.

Il en existe de nombreuses et en général, la distance choisie est celle de *Kullback-Leibler*. Il est fréquent d'inclure dans cette mesure de distance un terme de pénalité qui favorise les modèles simples ayant peu de relations causales. Notons enfin que si l'on souhaite utiliser cette procédure automatique, l'étape de spécification des lois conditionnelles des variables sachant leur(s) parent(s) est cruciale.

Il est en effet nécessaire de choisir des lois conditionnelles compatibles entre les différents modèles comparés. Dans le cas contraire, le choix automatique perd toute objectivité dans la mesure où il est énormément dépendant des lois conditionnelles choisies.

Conclusion

Nous nous sommes efforcés de donner dans cet article une introduction complète aux réseaux Bayésiens. Comme pour de nombreux outils récents et puissants de l'IA, entrer dans les détails demanderait un livre entier, à l'image de la référence [4], par exemple. On constate cependant que malgré des racines mathématiques profondes, les réseaux Bayésiens sont abordables relativement facilement.

De plus, pour des modèles simples comme celui de l'alarme, il n'est pas nécessaire de mettre en œuvre des algorithmes complexes pour exploiter le réseau. Des réseaux Bayésiens élémentaires sont d'ailleurs utilisés dans les jeux vidéo pour fournir des mécanismes d'IA simples et robustes.

Paul Tozour explique ainsi dans [15] comment donner des facultés déductives simples à un garde dans un jeu d'infiltration à la *Thief* : comme dans l'exemple de l'alarme, le garde considère deux variables d'intérêt, Voleur et Rat, qui sont les sources possibles des différents bruits perçus.

En plus de la variable d'information Bruit, le réseau peut contenir une variable Ombre pour la perception visuelle, ainsi que toutes sortes de variables d'information représentant l'environnement du garde. En fonction de l'état des variables d'information, le garde estime la probabilité de présence d'un voleur (le joueur !) ou d'un rat, puis décide de la conduite à tenir en s'appuyant sur les valeurs obtenues.

De même, l'exemple du filtrage de spam montre que la classification bayésienne naïve est relativement facile à mettre en œuvre, sans algorithmes évolués ni formation mathématique de haut niveau.

En fait, si on se limite à des graphes simples pour lesquels on fixe les lois de probabilité *a priori* ou simplement à partir d'observations expérimentales, les réseaux Bayésiens peuvent être intégrés sans difficulté dans toute application qui nécessite une forme d'intelligence.

Les vraies difficultés commencent quand on doit s'attaquer à des graphes volumineux ou pire encore, quand on ne connaît pas de modèle *a priori* et qu'on doit donc mettre en œuvre une méthode de construction automatique du réseau à partir des observations expérimentales.

Dans ces situations, les réseaux Bayésiens offrent des caractéristiques uniques et sont en fait l'un des seuls outils efficaces pour le raisonnement dans l'incertain. Par contre, ils demandent alors un double investissement, à la fois mathématique et algorithmique. C'est le cas de toutes les méthodes de l'IA de haut niveau.

Références

[1] Eric Lacombe et Fabrice Rossi - L'intelligence artificielle des jeux de stratégie classiques, GNU/Linux Magazine France, No 55, Novembre 2003.

[2] Fabrice Rossi - L'ordinateur peut-il lire dans votre esprit ?, GNU/Linux Magazine France, No 58, Février 2004.

[3] R.E. Neapolitan - Probabilistic Reasoning in Expert Systems, Wiley, New York, 1990.

[4] F.V. Jensen - An introduction to Bayesian networks, University College London Press, 1996.

[5] J. Pearl - Probabilistic Reasoning in Intelligent Systems, CA: Morgan Kauffman, San Mateo, 1988.

[6] Ian Hacking - L'émergence de la probabilité, Liber (Seuil), 2002.

[7] S.L. Lauritzen - Graphical Models, Clarendon Press, Oxford, 1996.

[8] G.F. Cooper - The computational complexity of probabilistic inference using Bayesian belief networks, Artificial Intelligence, 42, 393-405, 1990.

[9] J.S. Yedidia, W.T. Freeman et Y. Weiss - Understanding Belief Propagation and Its Generalizations, Mitsubishi Electric Research Laboratories, 2002.

[10] Une liste de logiciels pour l'exploitation des réseaux Bayésiens, maintenue par Kevin Murphy, est disponible à l'URL suivante : http://www.ai.mit.edu/~murphyk/Bayes/bnsoft.html

[11] Le package Grappa développé par Peter Green est disponible à l'URL suivante : http://www.stats.bris.ac.uk/~peter/Grappa/. Les pages de P. Green contiennent aussi des cours très intéressants sur les réseaux Bayésiens.

[12] Site du logiciel R : http://www.r-project.org

[13] Fabrice Rossi - Filtrage de SPAM par méthodes probabilistes, Multi-System & Internet Security Cookbook (MISC), No 7, Mai 2003.

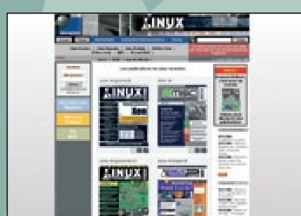
[14] Site du package Deal (GPL) : http://www.math.auc.dk/novo/deal/

[15] Paul Tozour - Introduction to Bayesian Networks and Reasoning Under Uncertainty, article du recueil AI Game Programming Wisdom, édité par Steve Rabin chez Charles River Media, 2002.

2 SITES INCONTOURNABLES

Toute l'actualité du magazine sur :

www.gnulinuxmag.com



Abonnements et anciens numéros en vente sur :

www.ed-diamond.com

Creative Commons

Paternité - Pas d'Utilisation Commerciale - Pas de Modification 2.0

Creative Commons n'est pas un cabinet d'avocats et ne fournit pas de services de conseil juridique. La distribution de la présente version de ce contrat ne crée aucune relation juridique entre les parties au contrat présenté ci-après et Creative Commons. Creative Commons fournit cette offre de contrat-type en l'état, à seule fin d'information. Creative Commons ne saurait être tenu responsable des éventuels préjudices résultant du contenu ou de l'utilisation de ce contrat.

Contrat

L'Oeuvre (telle que définie ci-dessous) est mise à disposition selon les termes du présent contrat appelé Contrat Public Creative Commons (dénommé ici « CPCC » ou « Contrat »). L'Oeuvre est protégée par le droit de la propriété littéraire et artistique (droit d'auteur, droits voisins, droits des producteurs de bases de données) ou toute autre loi applicable. Toute utilisation de l'Oeuvre autrement qu'explicitement autorisée selon ce Contrat ou le droit applicable est interdite.

L'exercice sur l'Oeuvre de tout droit proposé par le présent contrat vaut acceptation de celui-ci. Selon les termes et les obligations du présent contrat, la partie Offrante propose à la partie Acceptante l'exercice de certains droits présentés ci-après, et l'Acceptant en approuve les termes et conditions d'utilisation.

1. Définitions

a. **« Oeuvre »** : oeuvre de l'esprit protégeable par le droit de la propriété littéraire et artistique ou toute loi applicable et qui est mise à disposition selon les termes du présent Contrat.
b. **« Oeuvre dite Collective »** : une oeuvre dans laquelle l'oeuvre, dans sa forme intégrale et non modifiée, est assemblée en un ensemble collectif avec d'autres contributions qui constituent en elles-mêmes des oeuvres séparées et indépendantes. Constituent notamment des Oeuvres dites Collectives les publications périodiques, les anthologies ou les encyclopédies. Aux termes de la présente autorisation, une oeuvre qui constitue une Oeuvre dite Collective ne sera pas considérée comme une Oeuvre dite Dérivée (telle que définie ci-après).
c. **« Oeuvre dite Dérivée »** : une oeuvre créée soit à partir de l'Oeuvre seule, soit à partir de l'Oeuvre et d'autres oeuvres préexistantes. Constituent notamment des Oeuvres dites Dérivées les traductions, les arrangements musicaux, les adaptations théâtrales, littéraires ou cinématographiques, les enregistrements sonores, les reproductions par un art ou un procédé quelconque, les résumés, ou toute autre forme sous laquelle l'Oeuvre puisse être remaniée, modifiée, transformée ou adaptée, à l'exception d'une oeuvre qui constitue une Oeuvre dite Collective. Une Oeuvre dite Collective ne sera pas considérée comme une Oeuvre dite Dérivée aux termes du présent Contrat. Dans le cas où l'Oeuvre serait une composition musicale ou un enregistrement sonore, la synchronisation de l'oeuvre avec une image animée sera considérée comme une Oeuvre dite Dérivée pour les propos de ce Contrat.
d. **« Auteur original »** : la ou les personnes physiques qui ont créé l'Oeuvre.
e. **« Offrant »** : la ou les personne(s) physique(s) ou morale(s) qui proposent la mise à disposition de l'Oeuvre selon les termes du présent Contrat.
f. **« Acceptant »** : la personne physique ou morale qui accepte le présent contrat et exerce des droits sans en avoir violé les termes au préalable ou qui a reçu l'autorisation expresse de l'Offrant d'exercer des droits dans le cadre du présent contrat malgré une précédente violation de ce contrat.

2. Exceptions aux droits exclusifs. Aucune disposition de ce contrat n'a pour intention de réduire, limiter ou restreindre les prérogatives issues des exceptions aux droits, de l'épuisement des droits ou d'autres limitations aux droits exclusifs des ayants droit selon le droit de la propriété littéraire et artistique ou les autres lois applicables.

3. Autorisation. Soumis aux termes et conditions définis dans cette autorisation, et ceci pendant toute la durée de protection de l'Oeuvre par le droit de la propriété littéraire et artistique ou le droit applicable, l'Offrant accorde à l'Acceptant l'autorisation mondiale d'exercer à titre gratuit et non exclusif les droits suivants :

a. reproduire l'Oeuvre, incorporer l'Oeuvre dans une ou plusieurs Oeuvres dites Collectives et reproduire l'Oeuvre telle qu'incorporée dans lesdites Oeuvres dites Collectives;
b. distribuer des exemplaires ou enregistrements, présenter, représenter ou communiquer l'Oeuvre au public par tout procédé technique, y compris incorporée dans des Oeuvres Collectives;
c. lorsque l'Oeuvre est une base de données, extraire et réutiliser des parties substantielles de l'Oeuvre.

Les droits mentionnés ci-dessus peuvent être exercés sur tous les supports, médias, procédés techniques et formats. Les droits ci-dessus incluent le droit d'effectuer les modifications nécessaires techniquement à l'exercice des droits dans d'autres formats et procédés techniques. L'exercice de tous les droits qui ne sont pas expressément autorisés par l'Offrant ou dont il n'aurait pas la gestion demeure réservé, notamment les mécanismes de gestion collective obligatoire applicables décrits à l'article 4(d).

4. Restrictions. L'autorisation accordée par l'article 3 est expressément assujettie et limitée par le respect des restrictions suivantes :

a. L'Acceptant peut reproduire, distribuer, représenter ou communiquer au public l'Oeuvre y compris par voie numérique uniquement selon les termes de ce Contrat. L'Acceptant doit inclure une copie ou l'adresse Internet (Identifiant Uniforme de Ressource) du présent Contrat à toute reproduction ou enregistrement de l'Oeuvre que l'Acceptant distribue, représente ou communique au public y compris par voie numérique. L'Acceptant ne peut pas offrir ou imposer de conditions d'utilisation de l'Oeuvre qui altèrent ou restreignent les termes du présent Contrat ou l'exercice des droits qui y sont accordés au bénéficiaire. L'Acceptant ne peut pas céder de droits sur l'Oeuvre. L'Acceptant doit conserver intactes toutes les informations qui renvoient à ce Contrat et à l'exonération de responsabilité. L'Acceptant ne peut pas reproduire, distribuer, représenter ou communiquer au public l'Oeuvre, y compris par voie numérique, en utilisant une mesure technique de contrôle d'accès ou de contrôle d'utilisation qui serait contradictoire avec les termes de cet Accord contractuel. Les mentions ci-dessus s'appliquent à l'Oeuvre telle qu'incorporée dans une Oeuvre dite Collective, mais, en dehors de l'Oeuvre en elle-même, ne soumettent pas l'Oeuvre dite Collective, aux termes du présent Contrat. Si l'Acceptant crée une Oeuvre dite Collective, à la demande de tout Offrant, il devra, dans la mesure du possible, retirer de l'Oeuvre dite Collective toute référence au dit Offrant, comme demandé. Si l'Acceptant crée une Oeuvre dite Collective, à la demande de tout Auteur, il devra, dans la mesure du possible, retirer de l'Oeuvre dite Collective toute référence au dit Auteur, comme demandé.

b. L'Acceptant ne peut exercer aucun des droits conférés par l'article 3 avec l'intention ou l'objectif d'obtenir un profit commercial ou une compensation financière personnelle. L'échange de l'Oeuvre avec d'autres Oeuvres protégées par le droit de la propriété littéraire et artistique par le partage électronique de fichiers, ou par tout autre moyen, n'est pas considéré comme un échange avec l'intention ou l'objectif d'un profit commercial ou d'une compensation financière personnelle, dans la mesure où aucun paiement ou compensation financière n'intervient en relation avec l'échange d'Oeuvres protégées.
c. Si l'Acceptant reproduit, distribue, représente ou communique l'Oeuvre au public, y compris par voie numérique, il doit conserver intactes toutes les informations sur le régime des droits et en attribuer la paternité à l'Auteur Original, de manière raisonnable au regard au médium ou au moyen utilisé. Il doit communiquer le nom de l'Auteur Original ou son éventuel pseudonyme s'il est indiqué ; le titre de l'Oeuvre Originale s'il est indiqué ; dans la mesure du possible, l'adresse Internet ou Identifiant Uniforme de Ressource (URI), s'il existe, spécifié par l'Offrant comme associé à l'Oeuvre, à moins que cette adresse ne renvoie pas aux informations légales (paternité et conditions d'utilisation de l'Oeuvre). Ces obligations d'attribution de paternité doivent être exécutées de manière raisonnable. Cependant, dans le cas d'une Oeuvre dite Collective, ces informations doivent, au minimum, apparaître à la place et de manière aussi visible que celles à laquelle apparaissent les informations de même nature.
d. Dans le cas où une utilisation de l'Oeuvre serait soumise à un régime légal de gestion collective obligatoire, l'Offrant se réserve le droit exclusif de collecter ces redevances par l'intermédiaire de la société de perception et de répartition des droits compétente. Sont notamment concernés la radiodiffusion et la communication dans un lieu public de phonogrammes publiés à des fins de commerce, certains cas de retransmission par câble et satellite, la copie privée d'Oeuvres fixées sur phonogrammes ou vidéogrammes, la reproduction par reprographie.

5. Garantie et exonération de responsabilité

a. En mettant l'Oeuvre à la disposition du public selon les termes de ce Contrat, l'Offrant déclare de bonne foi qu'à sa connaissance et dans les limites d'une enquête raisonnable :
    i. L'Offrant a obtenu tous les droits sur l'Oeuvre nécessaires pour pouvoir autoriser l'exercice des droits accordés par le présent Contrat, et permettre la jouissance paisible et l'exercice licite de ces droits, ceci sans que l'Acceptant n'ait aucune obligation de verser de rémunération ou tout autre paiement ou droits, dans la limite des mécanismes de gestion collective obligatoire applicables décrits à l'article 4(e);
b. L'Oeuvre n'est constitutive ni d'une violation des droits de tiers, notamment du droit de la propriété littéraire et artistique, du droit des marques, du droit de l'information, du droit civil ou de tout autre droit, ni de diffamation, de violation de la vie privée ou de tout autre préjudice délictuel à l'égard de toute tierce partie.
c. A l'exception des situations expressément mentionnées dans le présent Contrat ou dans un autre accord écrit, ou exigées par la loi applicable, l'Oeuvre est mise à disposition en l'état sans garantie d'aucune sorte, qu'elle soit expresse ou tacite, y compris à l'égard du contenu ou de l'exactitude de l'Oeuvre.

6. Limitation de responsabilité. A l'exception des garanties d'ordre public imposées par la loi applicable et des réparations imposées par le régime de la responsabilité vis-à-vis d'un tiers en raison de la violation des garanties prévues par l'article 5 du présent contrat, l'Offrant ne sera en aucun cas tenu responsable vis-à-vis de l'Acceptant, sur la base d'aucune théorie légale ni en raison d'aucun préjudice direct, indirect, matériel ou moral, résultant de l'exécution du présent Contrat ou de l'utilisation de l'Oeuvre, y compris dans l'hypothèse où l'Offrant avait connaissance de la possible existence d'un tel préjudice.

7. Résiliation

a. Tout manquement aux termes du contrat par l'Acceptant entraîne la résiliation automatique du Contrat et la fin des droits qui en découlent. Cependant, le contrat conserve ses effets envers les personnes physiques ou morales qui ont reçu de la part de l'Acceptant, en exécution du présent contrat, la mise à disposition d'Oeuvres dites Dérivées, ou d'Oeuvres dites Collectives, ceci tant qu'elles respectent pleinement leurs obligations. Les sections 1, 2, 5, 6 et 7 du contrat continuent à s'appliquer après la résiliation de celui-ci.
b. Dans les limites indiquées ci-dessus, le présent Contrat s'applique pendant toute la durée de protection de l'Oeuvre selon le droit applicable. Néanmoins, l'Offrant se réserve à tout moment le droit d'exploiter l'Oeuvre sous des conditions contractuelles différentes, ou d'en cesser la diffusion; cependant, le recours à cette option ne doit pas conduire à retirer les effets du présent Contrat (ou de tout contrat qui a été ou doit être accordé selon les termes de ce Contrat), et ce Contrat continuera à s'appliquer dans tous ses effets jusqu'à ce que sa résiliation intervienne dans les conditions décrites ci-dessus.

8. Divers

a. A chaque reproduction ou communication au public par voie numérique de l'Oeuvre ou d'une Oeuvre dite Collective par l'Acceptant, l'Offrant propose au bénéficiaire une offre de mise à disposition de l'Oeuvre dans des termes et conditions identiques à ceux accordés à la partie Acceptante dans le présent Contrat.
b. La nullité ou l'inapplicabilité d'une quelconque disposition de ce Contrat au regard de la loi applicable n'affecte pas celle des autres dispositions qui resteront pleinement valides et applicables. Sans action additionnelle par les parties à cet accord, lesdites dispositions devront être interprétées dans la mesure minimum nécessaire à leur validité et leur applicabilité.
c. Aucune limite, renonciation ou modification des termes ou dispositions du présent Contrat ne pourra être acceptée sans le consentement écrit et signé de la partie compétente.
d. Ce Contrat constitue le seul accord entre les parties à propos de l'Oeuvre mise ici à disposition. Il n'existe aucun élément annexe, accord supplémentaire ou mandat portant sur cette Oeuvre en dehors des éléments mentionnés ici. L'Offrant ne sera tenu par aucune disposition supplémentaire qui pourrait apparaître dans une quelconque communication en provenance de l'Acceptant. Ce Contrat ne peut être modifié sans l'accord mutuel écrit de l'Offrant et de l'Acceptant.
e. Le droit applicable est le droit français.

Creative Commons n'est pas partie à ce Contrat et n'offre aucune forme de garantie relative à l'Oeuvre. Creative Commons décline toute responsabilité à l'égard de l'Acceptant ou de toute autre partie, quel que soit le fondement légal de cette responsabilité et quel que soit le préjudice subi, direct, indirect, matériel ou moral, qui surviendrait en rapport avec le présent Contrat. Cependant, si Creative Commons s'est expressément identifié comme Offrant pour mettre une Oeuvre à disposition selon les termes de ce Contrat, Creative Commons jouira de tous les droits et obligations d'un Offrant.

A l'exception des fins limitées à informer le public que l'Oeuvre est mise à disposition sous CPCC, aucune des parties n'utilisera la marque « Creative Commons » ou toute autre indication ou logo afférent sans le consentement préalable écrit de Creative Commons. Toute utilisation autorisée devra être effectuée en conformité avec les lignes directrices de Creative Commons à jour au moment de l'utilisation, telles qu'elles sont disponibles sur son site Internet ou sur simple demande.

Creative Commons peut être contacté à http://creativecommons.org/.