SûretéGlobale.Org
La Guitonnière
49770 La Meignanne

Téléphone : +33 241 777 886
Télécopie : +33 241 200 987
Portable : +33 6 83 01 01 80
Adresse de messagerie :
c.courtois@sureteglobale.org

APPORT DES RESEAUX BAYESIENS DANS LA PREVENTION DE LA DELINQUANCE

Une approche novatrice dans l'optimisation du ciblage et l'étude des interactions

Apport des réseaux bayésiens dans la prévention de la délinquance

Une approche novatrice dans l'optimisation du ciblage

" Une simple règle de l'arithmétique des probabilités donne un théorème exact qui modélise nos intuitions. La probabilité d'un évènement, sur l'affirmation qu'une hypothèse est vraie, est appelée la vraisemblance de l'hypothèse. Le théorème de Bayes affirme que la probabilité à postériori d'une hypothèse devrait être proportionnelle au produit de la probabilité à priori par la vraisemblance de l'hypothèse. C'est ainsi que l'on formalise l'idée d'apprentissage par expérience."

Hacking, I (2001). L'archéologue du probable. *Science et avenir. 128 :* 8-13.

UN EXEMPLE CONCRET : APPLICATION AUX DONNÉES DE L'ÉTAT 4001

En utilisant le logiciel BAYESIA CRIME ANALYST, développé par SûretéGlobale.Org en collaboration avec la société BAYESIA[1], qui offre des fonctionnalités évoluées (BAYESIA est le leader mondial de ce type de logiciel), nous avons introduit comme données observées l'état 4001 d'une ville de la région parisienne de janvier 2008 à mars 2008.

Ces données brutes se présentent sous cette forme sous Excel :

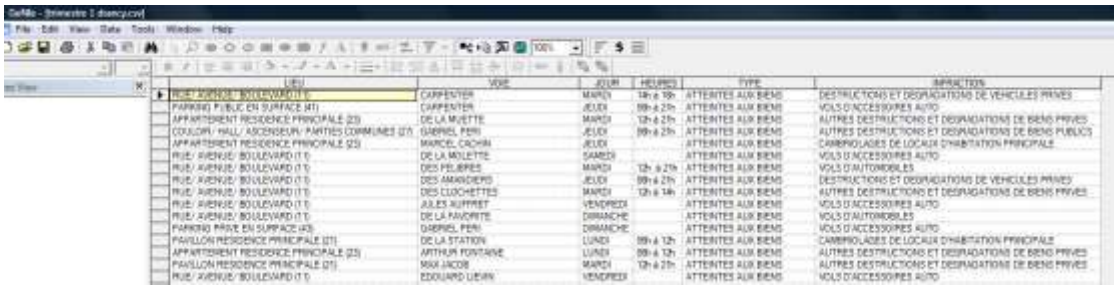
A2 — RUE/ AVENUE/ BOULEVARD (11)

	LIEU	VOIE	JOUR	HEURES	TYPE	INFRACTION
2	RUE/ AVENUE/ BOULEVARD (11)	CARPENTER	MARDI	14h à 16h	ATTEINTES AUX BIENS	DESTRUCTIONS ET DEGRADATIONS DE VEHICULES PRIVES
3	PARKING PUBLIC EN SURFACE (41)	CARPENTER	JEUDI	06h à 21h	ATTEINTES AUX BIENS	VOLS D'ACCESSOIRES AUTO
4	APPARTEMENT RESIDENCE PRINCIPALE (22)	DE LA MUETTE	MARDI	12h à 21h	ATTEINTES AUX BIENS	AUTRES DESTRUCTIONS ET DEGRADATIONS DE BIENS PRIVES
5	COULOIR/ HALL/ ASCENSEUR/ PARTIES COMMUNES (27)	GABRIEL PERI	JEUDI	06h à 21h	ATTEINTES AUX BIENS	AUTRES DESTRUCTIONS ET DEGRADATIONS DE BIENS PUBLICS
6	APPARTEMENT RESIDENCE PRINCIPALE (23)	MARCEL CACHIN	JEUDI		ATTEINTES AUX BIENS	CAMBRIOLAGES DE LOCAUX D'HABITATION PRINCIPALE
7	RUE/ AVENUE/ BOULEVARD (11)	DE LA MOLETTE	SAMEDI		ATTEINTES AUX BIENS	VOLS D'ACCESSOIRES AUTO
8	RUE/ AVENUE/ BOULEVARD (11)	DES FEUILLES	MARDI	12h à 21h	ATTEINTES AUX BIENS	VOLS D'AUTOMOBILES
9	RUE/ AVENUE/ BOULEVARD (11)	DES AMANDIERS	JEUDI	06h à 21h	ATTEINTES AUX BIENS	DESTRUCTIONS ET DEGRADATIONS DE VEHICULES PRIVES
10	RUE/ AVENUE/ BOULEVARD (11)	DES CLOCHETTES	MARDI	12h à 14h	ATTEINTES AUX BIENS	AUTRES DESTRUCTIONS ET DEGRADATIONS DE BIENS PRIVES
11	RUE/ AVENUE/ BOULEVARD (11)	JULES AUFFRET	VENDREDI		ATTEINTES AUX BIENS	VOLS D'ACCESSOIRES AUTO
12	RUE/ AVENUE/ BOULEVARD (11)	DE LA FAVORITE	DIMANCHE		ATTEINTES AUX BIENS	VOLS D'AUTOMOBILES

La base de test contient 484 lignes, correspondant à autant de plaintes enregistrées dans la période du 1er trimestre 2008 sur une commune de la région parisienne. Nous avons conservés les données d'infraction, mais utilisés des adresses d'une autre commune, pour des soucis de confidentialité.

Cette base à été exportée sous forme de fichiers .csv (séparateur point-virgule) puis importée dans importée dans BAYESIA CRIME ANALYST :

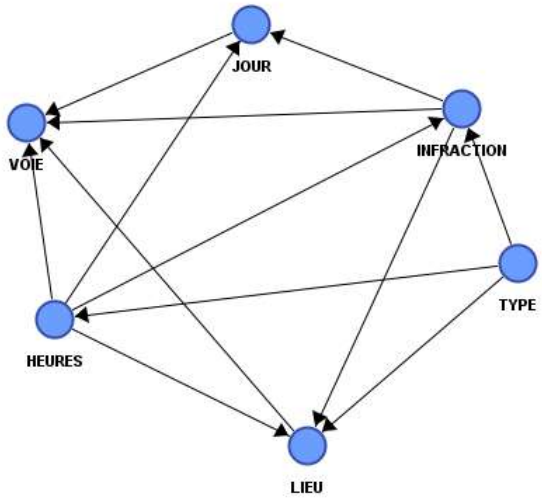
[1] Site web : www.bayesia.com



Nous avons ensuite complétés les données manquantes (heures de 00h à 24h quand elles ne sont pas renseignées) et regroupées des items orthographiés différemment pour s'assurer de la cohérence des données.

L'ensemble est ensuite « mouliné » par le logiciel, pour générer automatiquement la structure, réaliser les tableaux statistiques et découvrir les causalités des variables.

On obtient ainsi automatiquement le graphe bayésien suivant :



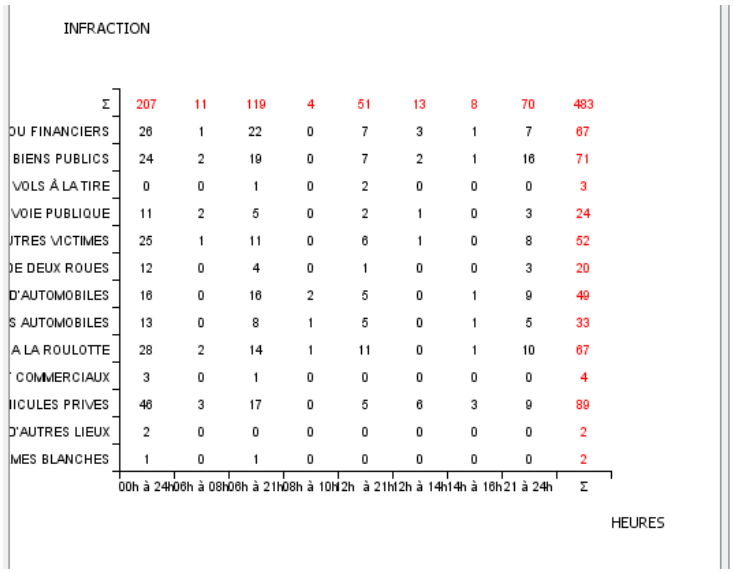
Chaque nœud comporte une table de probabilités, dépendante des données observées, et calculée par le logiciel :

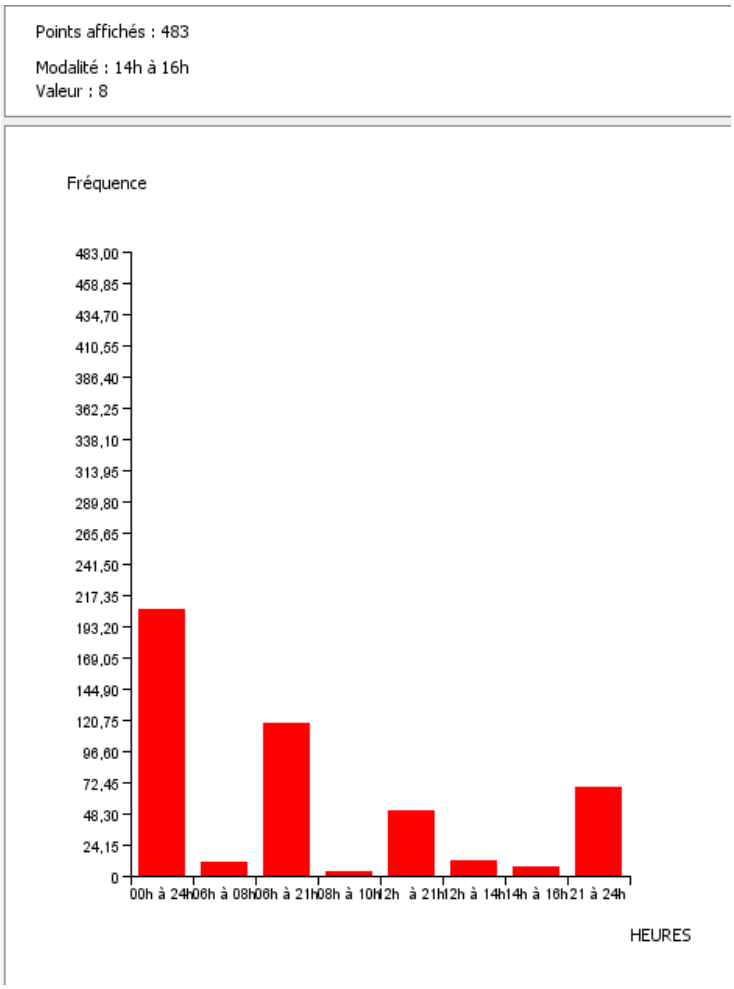
Mode de visualisation

Déterministe | Equation | Occurrences

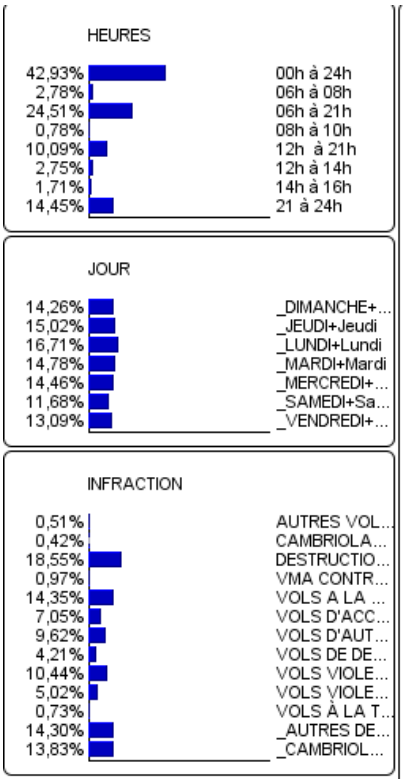
TYPE	00h à 24h	06h à 08h	06h à 21h	08h à 10h	12h à 21h	12h
ATTEINTES ...	30,702	0,000	29,825	0,000	12,281	
ATTEINTES ...	48,750	3,750	21,250	0,000	10,000	
CAMBRIOL...	44,000	2,000	30,000	0,000	10,000	
DESTRUCTI...	50,000	0,000	0,000	0,000	0,000	
DESTRUCTI...	0,000	0,000	0,000	0,000	0,000	
DESTRUCTI...	38,889	3,704	27,778	0,000	5,556	
DESTRUCTI...	53,968	4,762	19,048	0,000	6,349	
DESTRUCTI...	100,000	0,000	0,000	0,000	0,000	
VOLS A LA ...	0,000	0,000	50,000	0,000	50,000	
VOLS LIES ...	46,552	1,724	21,552	3,448	13,793	

Cette table peut aussi être vue sous forme de graphe statistique, ou de matrice d'occurrences :



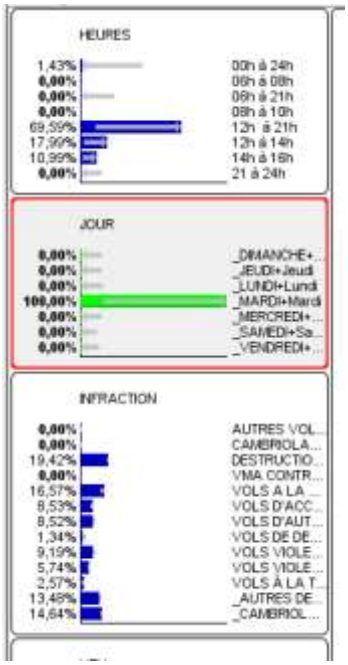


Le passage en mode propagation permet d'avoir les résultats qu'on obtient classiquement avec des logiciels d'analyses graphiques :



Que va-t-il se passer ce matin ?

Une première application directe de la propagation et de l'inférence est d'instancier une ou des variables (forcer une valeur de cette variable), par exemple le jour de la semaine et l'heure :



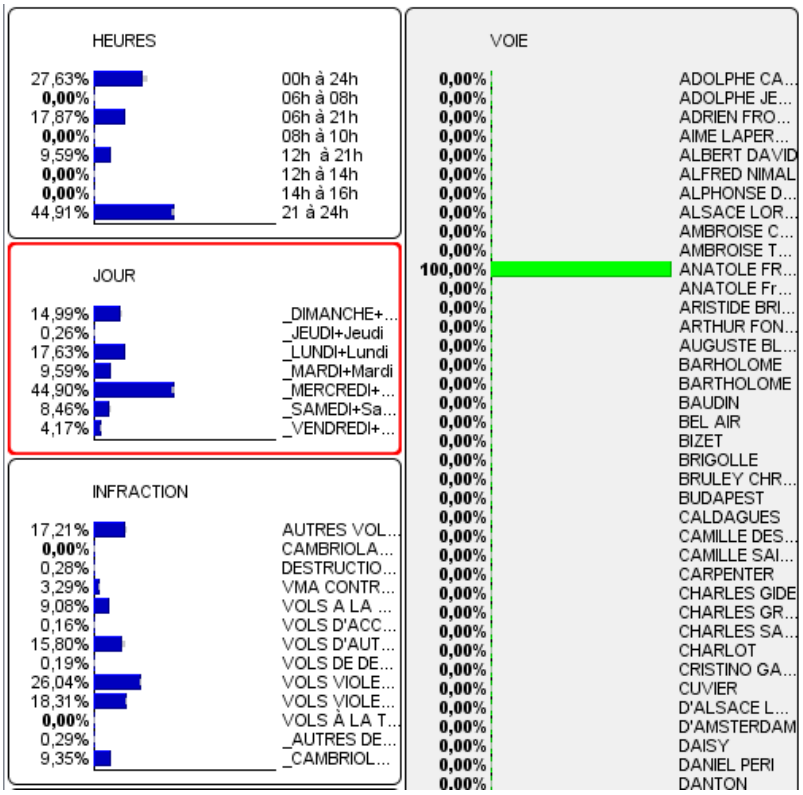
Nous sommes donc un mardi (forçage de la probabilité de la variable « jour » à 100% pour l'occurrence « mardi »).
Nous allons donc maintenant demander la propagation de ces probabilités de ces variables observées vers les autres variables.

Le mardi, les infractions ont surtout lieu de 12h00 à 21h00 (70%), dans la rue (64%), pour des dégradations (19% de dégradations de véhicules, 10% autres) des vols à la roulottes (16%) et des vols violents (10%).

En poussant un peu plus loin, par exemple en cherchant pour chaque jour de la semaine, on se rend compte que prés de 10% des délits commis le dimanche le sont rue « Henri Barbusse » sur la période considérée.

Que se passe-t-il là bas ?

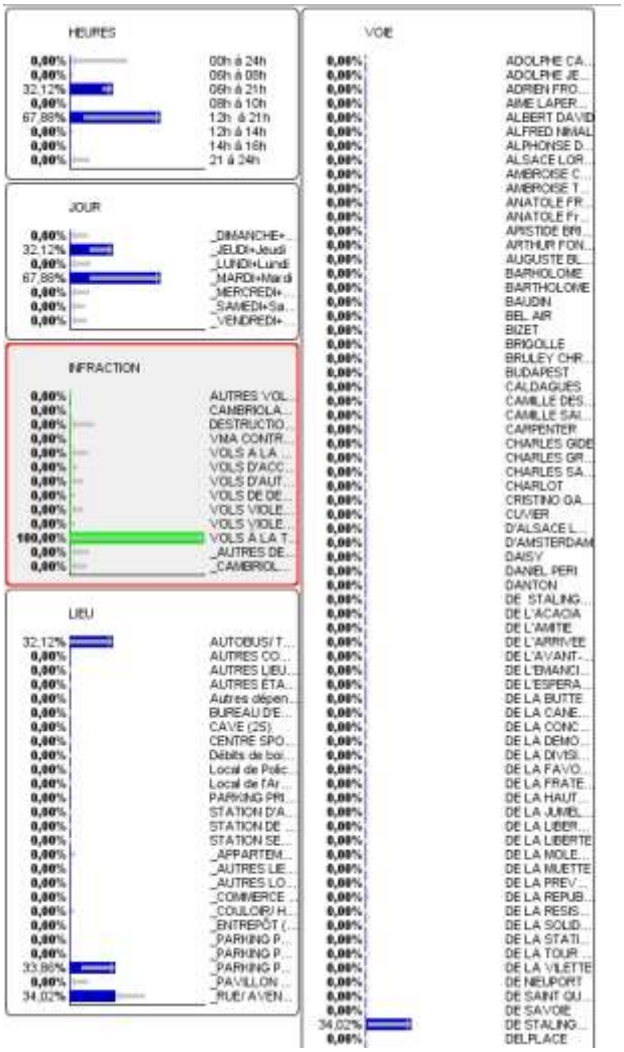
A l'inverse, le forçage d'un lieu géographique permet de se rendre compte de la criminalité sur l'endroit : en instanciant la rue « Anatole France », on se rend compte :



- Que de nombreux délits sont commis entre 21h et 24h00 (45%)
- Qu'il s'agit essentiellement de vols violents (26% sans armes, 18% avec armes blanches)
- Que c'est le mercredi où les occurrences sont les plus nombreuses (45%) puis le lundi et le dimanche.

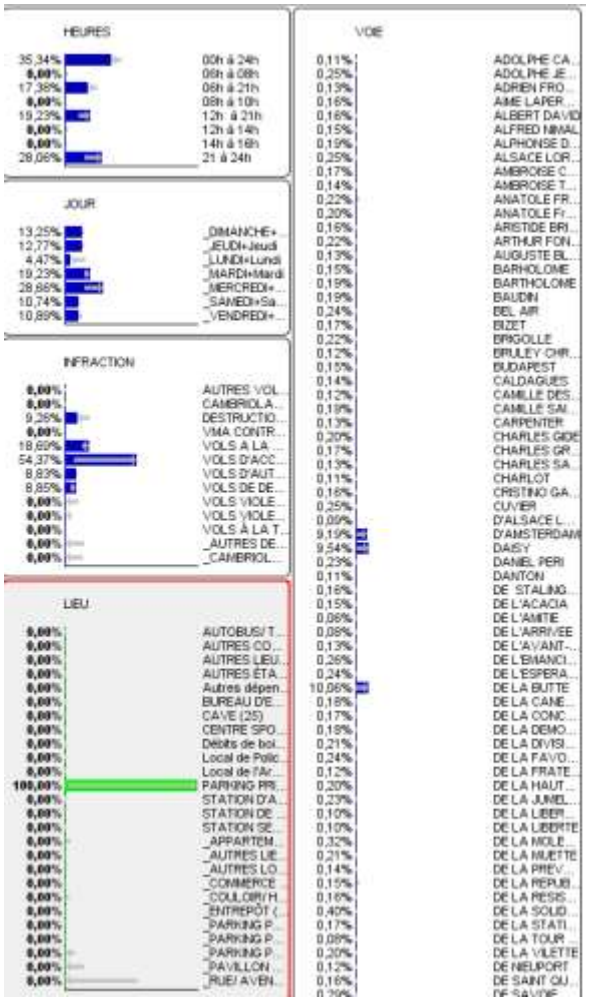
Où et quand volent-on à la tire ?

En instanciant cette fois le type de délits, on obtient :



C'est le mardi (68%) et le jeudi (32%) de 12h à 21h00 (68%), dans les rues « Jean Varnet », « Stalingrad », « Saint Stenay » (34% chacune), dans les transports en commun, les parkings et dans la rue (environ 33% chacun).

Que se passe-t-il dans les parkings souterrains ?



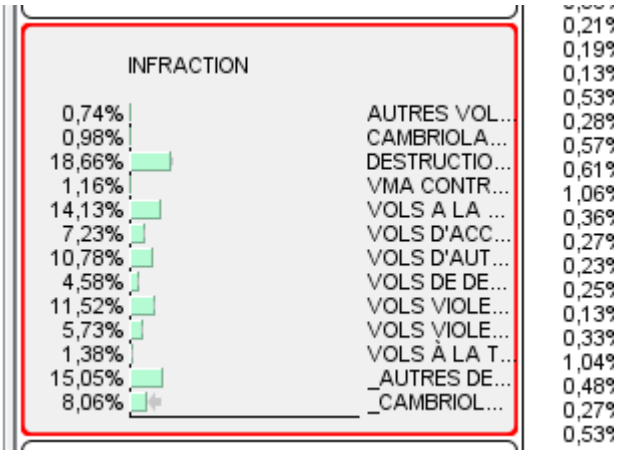
Il s'agit essentiellement de vols de deux roues, et de vols à la roulotte commis le mercredi, le vendredi et le samedi principalement dans les rues « Paul Vaillant Couturier », « Pierre et Marie Curie » et « Saint Stenay ».

On le voit, les applications de l'inférence bayésienne sont multiples, et plus on a de données, mieux ça vaut. Les réseaux bayésiens sont capables d'apprendre, et de propager des probabilités avec des intervalles de confiance qui croissent avec le nombre d'observations, et ce en un temps record.

La présentation de ces données, la facilité avec laquelle on peut simuler des situations permet de découvrir rapidement des éléments clés d'analyse. Si certains résultats pourraient être obtenus avec des analyses classiques, leurs découvertes prendraient énormément de temps.

Mais que se passerait –il s'il y avait moins de cambriolages ?

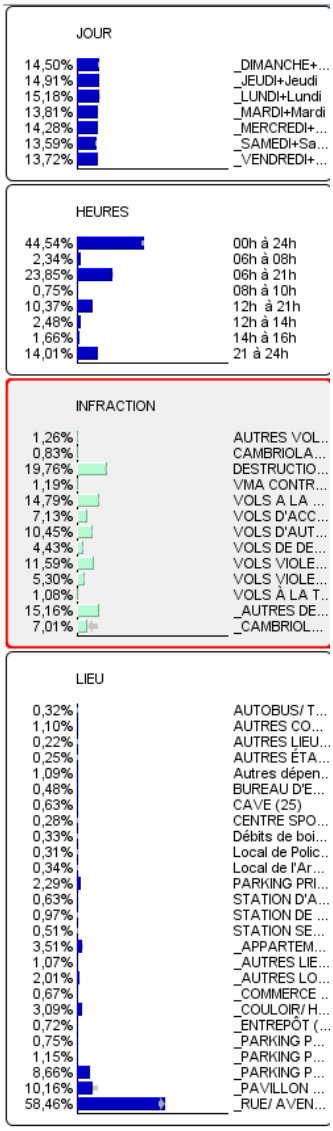
Un autre élément important en faveur des réseaux bayésiens est la possibilité de simuler des situations, en jouant directement sur les données.
Nous avons simulé une baisse des niveaux de cambriolages (-50%) :



Regardons ce qu'il se passerait à partir de là :

Avant :

Après :



On note que dans le cas d'une baisse des cambriolages, les probabilités sont que le vol à la roulotte augmenterait en proportion, et que les destructions et dégradations de véhicules augmenteraient. Globalement, la délinquance baisserait sur tout le territoire, avec une déportation des faits vers les délinquance de rue (dégradations, violences..)

Il ne s'agit pas là d'une nouvelle répartition sur 100%, mais bien des résultats des causes à effets calculés par le logiciel en fonction des données observées. Ce sont les règles (équations) déterminées par le calcul entre les différentes variables qui ont permis de recalculer, en fonction des modifications effectuées, les incidences probables de celles-ci sur l'ensemble des variables restantes.

Ces différents exemples montrent ce qu'il est possible de faire simplement avec des outils de type réseau bayésiens. Il est encore possible d'aller plus loin : en introduisant la notion de coût ou d'utilité, en pondérant les chiffres par la population ou les flux, en déclinant les notions de zones ou de quartiers, voire de latitude/longitude pour obtenir des cartes, en ajoutant des états permettant de mettre en œuvre des fonctions de diagnostics automatiques, ou des tests itératifs permettant de jouer des scénarios.

Évidemment, ce type de traitement peut trouver de nombreuses applications en matière de sécurité : données pompiers, police municipale, sécurité routière, bailleurs, données sociales, autant d'éléments qui peuvent être ajoutés, dans le cadre d'un CLS par exemple, pour mieux analyser une situation, optimiser les ressources, et arriver finalement à une analyse quasi-prédictive. De quoi optimiser l'action des forces, ou comprendre en détail des phénomènes.

Pour peu qu'on puisse par ailleurs traiter plus de données (enquête de victimation par exemple) ou qu'il soit possible de traiter aussi les données victimes et auteurs (STIC), on pourrait constituer un outil décisionnel et opérationnel dont la portée serait immédiate sur la prévention et la répression de la délinquance.

On le voit, à partir de données simples voire même triviales, les capacités des réseaux bayésiens vont bien au-delà de tout ce qui a pu se faire en matière d'analyse de la délinquance.

COMMENT ÇA MARCHE - UN PEU DE MATHS…

Technique mathématique combinant statistiques et intelligence artificielle, les réseaux bayésiens permettent d'analyser de grandes quantités de données pour en extraire des connaissances utiles à la prise de décision, contrôler ou prévoir le comportement d'un système, diagnostiquer les causes d'un phénomène, etc.

Les réseaux bayésiens, qui doivent leur nom aux travaux de Thomas Bayes au XVIIIe siècle sur la théorie des probabilités, sont le résultat de recherches effectuées dans les années 1980, dues à J. Pearl à UCLA et à une équipe de recherche danoise à l'université de Aalborg.

Les réseaux bayésiens sont utilisés dans de nombreux domaines : santé (diagnostic, localisation de gènes), industrie (contrôle d'automates ou de robots), informatique et réseaux (agents intelligents), marketing (data mining, gestion de la relation client), banque et finances (scoring, analyse financière), management (aide à la décision, knowledge management, gestion du risque), etc.

Les réseaux bayésiens, initiés par Judea Pearl dés les années 80, sont des modèles graphiques qui représentent les relations probabilisées entre un ensemble de variables. Il s'agit donc d'une approche probabiliste, ou la valeur de chaque variable est induite – ou non – par les valeurs des autres variables liées.

L'idée essentielle est donc de calculer la probabilité d'un événement donné en fonction de la probabilité d'autres événements préalablement observés. Il s'agit de déduire des effets à partir des causes.

Un réseau bayésien est donc basé sur cette question : Un événement A s'est produit. Quelle est la probabilité que ce soit la cause Mi qui l'ait produit ?

$P(Mi\ |A) = \frac{P(A|Mi\ ) \times P(Mi\ )}{P(A)}$

Avec P(Mi|A) : probabilité a posteriori
Et P(A) : constante (pour chaque Mi)

Le théorème de Bayes généralisé donne :

P(A1 . . . An) = P(A1)P(A2|A1)P(A3|A1, A2) . . . P(An|A1 . . . An−1)

C'est ce théorème qui est mis en œuvre dans les logiciels de réseaux bayèsiens.

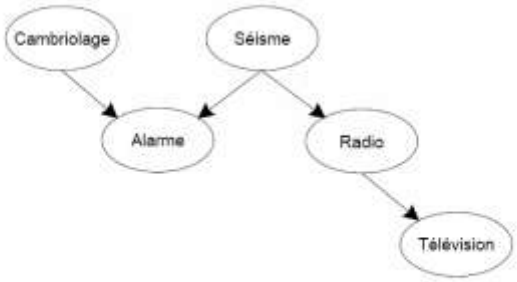
Un réseau bayésien B = (G; θ) est donc défini par :

- L'ensemble des variables aléatoires observables X = $\{X|X_n\}$
- G = (X;E), graphe dirigé sans circuit (DAG), ou chaque nœud est associe a une variable de X.
- $\theta = \{\theta_i\} = \{P(Xi|Pa(Xi))\}$ ensemble des distributions de probabilités de chaque nœud Xi, qui conditionnellement a ses parents immédiats dans le graphe G.

Le graphe d'un réseau bayésiens permet ainsi de représenter d'une façon visuelle les relations (dépendances et indépendances) entre les variables du système. Les distributions de probabilités permettent d'enrichir cette structure graphique par une quantification de ces relations. Les probabilités dans un réseau bayésiens permettent de représenter l'aspect incertain qui relie les variables.

La figure 1.1 est un exemple de réseau bayésiens modélisant la probabilité de déclencher une alarme suite aux deux causes possibles : cambriolage et séisme. Il modélise aussi la probabilité que la radio et la télévision annoncent un séisme sachant l'état vrai ou faux (incertain) de la variable "séisme".

Cet exemple modélise les relations qui relient les variables : le cambriolage et le séisme provoquent le déclenchement de l'alarme, un séisme entraine l'annonce de cet événement à la radio et par la suite a la télévision...

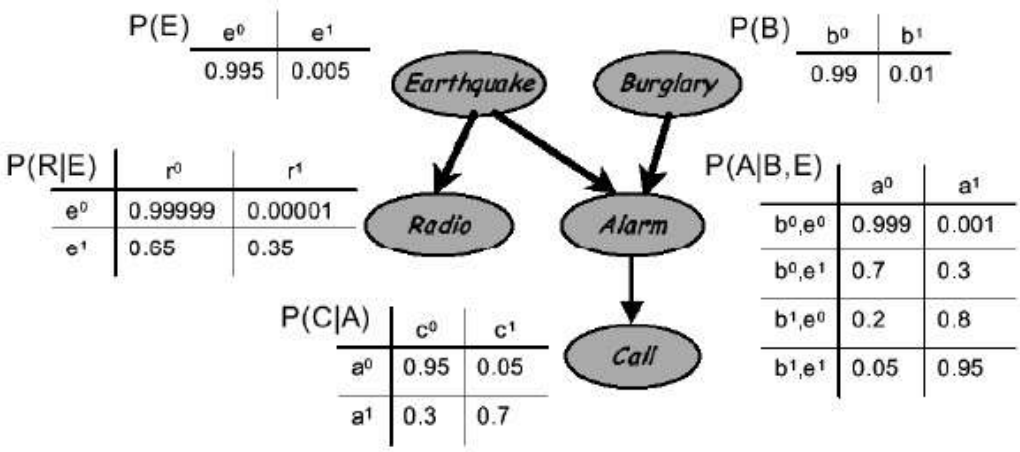


TAB. 1.1 – Exemple de réseau bayésien

Les réseaux bayésiens permettent d'autre part de représenter de manière compacte la distribution de probabilité jointe sur l'ensemble des variables.

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i|Pa(X_i))$$

A chaque nœud est donc associée une table de probabilité, de 0 (impossible) à 1 (vrai, observé) :



Cette décomposition de la loi jointe permet de faire des réseaux bayésiens un modèle économique pour représenter des distributions de probabilités. Ces simplifications apportées par cette loi jointe ont permis de rendre possible l'apparition des algorithmes d'inférence.

Ces algorithmes permettent, a partir de la structure d'un réseau bayésiens et des distributions des probabilités associée, de calculer la probabilité de n'importe quelle variable du modèle a partir de l'observation même partielle des autres variables.

L'inférence dans un réseau bayésien est un calcul de probabilités conditionnelles : elle consiste en une mise à jour des probabilités des variables non observées après observation des valeurs d'un certain nombre d'autres variables.

Autrement dit, dans l'exemple ci-dessus, si le téléphone sonne, il y tant de probabilité qu'il y ait un séisme, ou un cambriolage, récursivement à l'observation de la variable téléphone sonne=vrai.

Cette propagation des probabilités permet aussi de simuler des conditions en « forçant » une variable à une valeur donnée (par exemple, « séisme= vrai).

Il s'agit de **l'instanciation** de variables, permettant le calcul des inférences.

Enfin, les algorithmes d'apprentissages, bien que complexes et nombreux, permettent, à partir des données observées, de créer le réseau bayésien correspondant, avec les tables de probabilités observées, et ce de façon automatique, même si les observations sont incomplètes.

L'une des utilisations les plus intéressantes des réseaux bayésiens est celle précisément de la causal knowledge discovery, c'est-à-dire de la recherche automatique des liens causaux entre les variables d'une base de données. À partir d'un ensemble de variables observées, même de façon lacunaire et incertaine, sur un certain nombre de cas indépendants, de puissants algorithmes de génération de réseaux bayésiens permettent de déduire en probabilité des liens de cause à effet entre les variables. Les algorithmes les plus complets peuvent apprendre, et ainsi reproduire, tant la structure que les paramètres du réseau.

C'est donc un outil de découverte de connaissances à partir de données, même incomplètes. Les réseaux bayésiens, et plus généralement les modèles graphiques sont des outils généraux développés assez récemment (Pearl 1988, Jensen 1996, Whittaker 1990, Jordan 1998).

Les RB sont un mode intéressant de représentation d'une structure d'influence entre divers faits, états ou hypothèses d'états. Ils **permettent de modéliser des informations telles que des dépendances causales, spatiales, temporelles, même si elles sont imparfaites ou manquantes.**

En fait, un réseau bayésien permet la transformation d'observations incomplètes en connaissance de la façon suivante :