



HAL
open science

Utilisation des Réseaux Bayésiens Dynamiques pour la prédiction de survenue des Infections Nosocomiales

Ghada Trabelsi, Mounir Ben Ayed, Adel M. Alimi

► **To cite this version:**

Ghada Trabelsi, Mounir Ben Ayed, Adel M. Alimi. Utilisation des Réseaux Bayésiens Dynamiques pour la prédiction de survenue des Infections Nosocomiales. 5èmes Journées Francophones sur les Réseaux Bayésiens (JFRB2010), May 2010, Nantes, France. hal-00467554

HAL Id: hal-00467554

<https://hal.science/hal-00467554>

Submitted on 28 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Utilisation des Réseaux Bayésiens Dynamiques pour la prédiction de survenue des Infections Nosocomiales

Ghada Trabelsi* — **Mounir Ben Ayed*** — **Adel M. Alimi***

* *REGIM: Research Group on Intelligent Machines*
Université de Sfax, Ecole Nationale des Ingénieurs de Sfax (ENIS)
Route de Soukra km 4 B.P.3038 Sfax, Tunisie
ghada.trabelsi@gmail.com , mounir.benayed@ieee.org, Adel.Alimi@ieee.org

RÉSUMÉ. Le travail se situe dans le domaine de l'extraction des connaissances à partir des données (ECD). L'ECD permet d'acquérir des informations pertinentes pour les systèmes interactifs d'aide à la décision (SIAD). Mais, dans plusieurs domaines, les données évoluent d'une manière dynamique et finissent par dépendre de plusieurs dimensions. Dans ce travail nous proposons une approche pour l'ECD à partir des données temporelles. Le Réseau Bayésien RB est une des techniques les plus utilisées en ECD. Pour cela, nous avons développé un SIAD basé sur les RB Dynamiques RBD. Notre objectif revient à fixer les meilleurs modèles de connaissances extraites par les RBD et à l'utiliser dans la prise de décision d'une manière dynamique. Une évaluation réalisée dans le service de réanimation de l'hôpital Habib Bourguiba à Sfax pour la lutte contre l'infection nosocomiale a montré son efficacité.

ABSTRACT. This Work is located in the domain of the knowledge Discovery from data (KDD). KDD permits to acquire applicable information for the the Decision Support Systems (DSS). But, in several domains, the data evolve with a dynamic manner and finish by depending on several dimensions. In this work we propose an approach for the KDD use the temporal data. The Bayesian Network BN is one of the techniques used in KDD. For it, we developed a KDD based on the Dynamic BN (DBN). Our objective comes back to fix the best knowledge models extracted by the DBN and to use it in the dynamic decision. An assessment achieved in the ICU of the Habib Bourguiba hospital in Sfax for struggle against the Nosocomial Infection showed its efficiency.

MOTS-CLÉS : Les Réseaux Bayésiens Dynamiques, les données temporelles, Les infections Nosocomiales

KEYWORDS: Dynamic Bayesian Networks, Temporal data, Nosocomial infections

1. Introduction

Plusieurs systèmes interactifs d'aide à la décision sont souvent basés sur le processus ECD. L'ECD contient plusieurs phases de traitement parmi les quelles la phase de fouille de données (Fayyad, 1996). La plus part des techniques de fouille de données utilisent des algorithmes qui fonctionnent sur des données fixes non temporelles (chaque individu est représenté par un vecteur de caractéristiques) (Lefébure *et al.*, 2001). Dans plusieurs domaines, telque celui de la santé, les données concernant un patient sont saisies à des moments différents plus ou moins périodiques. Ainsi, le temps est un important concept dans les soins médicaux (Shahar, 2000), (Peelen *et al.*, 2010). L'exploitation de ces données (que certains qualifient de complexes) est difficile et délicate. Peu des travaux proposent des algorithmes pour l'extraction des connaissances à partir de données multidimensionnelles et temporelles.

D'après Becker *et al.*, l'information n'est pas la connaissance (Becker *et al.*, 1999). Certaines de ces connaissances peuvent être extraites à l'aide d'un outil décisionnel qu'est l'Extraction de Connaissances à partir des Données (ECD) (Fayyad, 1996). A mesure que se développent les technologies permettant de stocker, d'échanger de l'information et d'y accéder, la question de l'analyse et de synthèse de ces informations devient essentielle. L'une des techniques intelligentes pour la fouille de données est les Réseaux Bayésiens. Nous sommes intéressés dans cet article par cette dernière technique. Dans ce cadre, plusieurs équipes ont étudié les Réseaux Bayésiens. Certaines d'entre elles ont étudié cette technique côté apprentissage et inférence (Leray, 2006), (Naim *et al.*, 2007) et d'autres l'ont étudiée en la comparant avec d'autres techniques de data mining comme le réseau de neurone (Correa *et al.*, 2009), KPPV (Pernkopf, 2005), etc. La plus part des études sont limitées sur les Réseaux Bayésiens statique (Antonucci *et al.*, 2008), (Correa *et al.*, 2009). Ce qui est insuffisant pour la prise de décision dynamique tel que le domaine sur le quel nous travaillons. Le contexte applicatif de notre projet est la lutte contre les infections nosocomiales (IN) des patients hospitalisés dans le service de réanimation du CHU Habib Bourguiba de Sfax (Tunisie).

Les données dont nous disposons sont temporelles et multidimensionnelles. La décision à prendre dépend de l'état actuel du patient et de ses observations antérieurs. Nous avons donc utilisé les Réseaux Bayésiens dynamiques comme technique d'analyse pour obtenir des modèles de connaissances qui évoluent avec le temps.

Le présent article est organisé en quatre sections. Dans la section 2, nous présenterons l'état de l'art en particulier les systèmes interactifs d'aide à la décision (SIAD) et la technique d'apprentissages adoptée. Nous expliquons l'aspect dynamique du contexte de notre travail qui est la lutte contre les infections nosocomiales dans la section 3. Dans cette section, nous décrivons aussi la réalisation d'un processus ECD comme outil pour l'aide à la prise de décision. Dans la section 4, nous discuterons les résultats obtenus par notre système. Finalement, une conclusion et plusieurs perspectives seront soulignées.

2. L'état de l'art

2.1. Les systèmes Interactifs d'Aide à la Décision (SIAD)

Le concept de Systèmes Interactifs d'Aide à la Décision (SIAD) a tout d'abord été introduit dans les années 70. Il est la traduction du concept de Decision Support Systems (DSSs). Un des tous premiers auteurs à l'avoir introduit et diffusé est Scott Morton (Scott, 1971) qui a introduit la notion de Systèmes de décision et de Gestion (Management Decision Systems).

(Sprague *et al.*, 1982) ont donné une définition très proche de celle proposée par Keen et Morton (Keen *et al.*, 1978). Les SIAD peuvent être caractérisés comme des systèmes informatisés, interactifs qui aident les décideurs en utilisant des données et des modèles pour résoudre des problèmes mal structurés. Aussi Lepreux a définie un SIAD comme un système interactif, qui doit assister un décideur tout au long de son processus de décision par des interactions adaptées. Il est composé d'outils de mesure, d'analyse, de comparaison qui doivent l'aider dans l'évaluation des solutions possibles. Les interactions entre l'utilisateur, le SIAD et l'ensemble des outils permettent à l'utilisateur de prendre une décision. Pour cela, les interactions entre le SIAD et l'utilisateur doivent respecter les processus de prise de décision de l'utilisateur (Lepreux, 2005). Certaines décisions nécessitent l'accès à des connaissances plus approfondies que d'autres. Il y a donc un lien entre les systèmes d'aide à la décision et les systèmes de connaissance (Lepreux, 2005). De plus, dans la littérature, on trouve des études (Courtney, 2001) qui traitent conjointement ces deux domaines et présentent la prise de décision basée sur la connaissance. Nous verrons dans la suite que nous avons pu utiliser la fouille de données (par l'extraction des modèles de connaissances à partir des données temporelles) pour aider l'expert à prendre une décision.

2.2. Les Réseaux Bayésiens Dynamiques

Les modèles graphiques probabilistes, et plus précisément les Réseaux Bayésiens RBs sont initiés par Judea Pearl dans les années 1980 (Pearl, 1988). Ils facilitent les systèmes de prédiction et d'abduction en Intelligence Artificielle. Les RBs forment un ensemble de modèles probabilistes pour de larges collections de variables. Ils sont couramment utilisés en intelligence artificielle et en apprentissage automatique (machine learning) (Jordan, 2004). Les RB ont pour objectif d'acquérir (apprendre), représenter et utiliser la connaissance.

Les RB ne modélisent pas des relations temporelles entre les variables. Ils représentent seulement des rapports probabilistes dans un ensemble de variables à un point de temps (Neapolitan, 2004). Dans plusieurs domaines, la modélisation des relations temporelles est très importante. Par exemple, dans le domaine de la médecine la prise de décision est dynamique (quotidienne). Nous allons expliquer, dans la suite, la technique des réseaux Bayésien dynamiques, avec la quel nous pouvons résoudre les problèmes de cet aspect.

Définition : Neapolitan a définie un RBD comme une extension des RB qui se

modélise sur un processus temporel (Neapolitan, 2004). Murphy l'a définie comme un RB spécial qui est utilisé pour les modèles de processus dynamique stochastique (Murphy, 2002).

Notation : Dans ce qui suit X_t^i définit une variable aléatoire continue ou discrète. Des valeurs de la variable aléatoire seront indiquées par les lettres minuscules telles que x_t^i . Pour une variable discrète qui prend r valeurs, $(x_k^i)^t$ définir une tâche spécifique pour $1 \leq k \leq r$, $1 \leq i \leq n$ et $1 \leq t \leq T$. Un ensemble de variables est nommé en caractères majuscules des lettres $X = X_1, X_2, \dots, X_n$.

On peut définir un RBD comme un RB qui contient T vecteurs aléatoires de variables $X[t] = X_1[t], \dots, X_n[t]$, un RB est caractérisé par :

1) Un RB initial consiste à i) un GAD ou DAG initial G_0 qui contient les variables dans $X[0]$; et ii) une initiale distribution de probabilité P_0 de ces variables.

2) Une transition de RB est caractérisée par i) une transition DAG $G_{->}$ qui contient les variables $X[t]$ et $X[t+1]$; et ii) une distribution de probabilité pour une transition $P_{->}$ avec la quelle on assigne une probabilité conditionnelle à chaque valeur donnée de $X[t+1]$ par chaque valeur de $X[t]$. Cette probabilité, est calculée pour chaque valeur $x[t+1]$ de $X[t+1]$ et valeur $x[t]$ de $X[t]$.

$$P_{->}(X[t+1] = x[t+1] | X[t] = x[t]) \quad [1]$$

3) Un RBD contient les variables de T vecteurs de variables aléatoires i) un DAG se compose d'un DAG G_0 et pour $0 \leq t \leq T-1$ un DAG $G_{->}$ évalué pour t ; et ii) la distribution de probabilité jointe est :

$$P(X[0], \dots, X[T]) = P_0(X[0]) \prod_{t=0}^{T-1} P_{->}(X[t+1] | X[t]) \quad [2]$$

3. Mise en place d'un RBD pour un SIAD basé sur l'ECD

3.1. Contexte applicatif

L'aspect applicatif de notre projet entre dans le cadre d'un grand projet visant à lutter contre les infections nosocomiales. Les infections nosocomiales (IN) représentent un des problèmes majeurs de santé publique. Une infection est considérée comme telle lorsqu'elle était absente au moment de l'admission du patient (Garner *et al.*, 1988). Lorsque l'état infectieux du patient à l'admission est inconnu, l'infection est classiquement considérée comme nosocomiale si elle apparaît après un délai de 48 heures d'hospitalisation. Plusieurs travaux ont proposé des systèmes informatiques de surveillance des IN basés sur les techniques de fouille de données (Brossette *et al.*, 2000).

Une étude sur la prévalence de la survenue d'infections nosocomiales dans le CHU Habib Bourguiba à Sfax, Tunisie (Kallel *et al.*, 2005), a montré que 17,9 % des 280

patients hospitalisés dans l'ensemble de l'hôpital ont été victimes d'une IN entre le 17 avril 2002 (minuit) et le 18 avril 2002 (minuit). Notre projet est en lien avec les médecins du service de réanimation de cet hôpital. Ces médecins sont les experts et futurs utilisateurs du SIAD que nous étudions et développons.

3.2. Description des données

On dispose d'une base de données médicale qui contient toutes les informations concernant les patients d'un service de réanimation à l'hôpital. Notre objectif est d'extraire les variables qui représenteront les noeuds pour notre modèle de connaissance statique : les variables utilisées dans notre modélisation sont :

- âge, poids : champs numériques, correspondants à l'âge et au poids des patients.
- date entreRea, date entrée hôpital, date sortie Rea : champs dates, correspond aux date d'entrée au service de réanimation, date d'entrée à l'hôpital et date de sortie du service de réanimation.
- origine : correspond à l'origine du patient, hôpital ou domicile.
- servOrig : service d'origine : Maternité, Pédiatrie, Neuro chirurgie...
- priseAnt d'ATB : dans ce champ est indiqué si le patient a pris des antibiotique, de valeurs binaires.
- catégorie : la catégorie de chaque patient soit Médical, Chirurgical ou Traumatique.
- cissue : ce champ est utilisé pour indiquer l'état du patient actuel : survécu ou décédé.
- infNos : ce champ indique si le patient a acquis une IN ou non.
- diag1, diag2 : ces deux champs représentent les diagnostics.

Nous avons décrit les champs des données fixes. Se sont des données dont la valeur ne change pas pendant la période d'hospitalisation d'un patient. Elles contiennent plusieurs informations hétérogènes, où nous avons 17 champs et 280 enregistrements (patients). Donc nous allons tout d'abord traiter cette base de données afin de réduire le nombre des variables et de ne laisser que des variables utiles à la prédiction. Nous rappelons que nous avons d'autre type de données. Elles sont caractérisées par le temps, elles ont une structure plus complexe que les données fixes. ce sont des données qui contiennent une valeur pour chaque série de temps (jour) pendant la période d'hospitalisation d'un patient.

Nous avons utilisé des tables qui contiennent des données temporelles (examInf, VerifierActe, avoirAnt). Ces données ont une structure très complexe (la représentation est multidimensionnelle) et engendrent des problèmes lors de la phase d'apprentissage dans les RB.

- Acte : sont les actes effectués pour un patient à une date donnée par (date_acte)

- examInf : les examens infectieux effectués pour un patient à une date donnée par (date_exam)
- Pres_Ant : concernant la consommation de l'antibiotique pour un patient à une date donnée par (date_Prise) et une dose précise.

Ces données sont irrégulières et bruitées. Mais elles contiennent des informations très pertinentes pour notre prédiction et classification. Comme nous l'avons mentionné, plus haut, nous disposons d'une base contenant 280 cas. Il existe dans Les trois tables plus que 5600 enregistrements (pour tous les patients de la base) concernant les observations saisies quotidiennement.

3.3. Modélisation

3.3.1. Construction du modèle de connaissance des données fixes (RB Statique)

Il existe deux méthodes de construction de structure, une qui se base sur les indépendances conditionnelles entre les variables (algorithme PC), et une autre basée sur les scores (algorithme MWST, GS et K2). Nous avons adopté ces algorithmes pour l'apprentissage des modèles statiques.

Notre construction des modèles est basée sur la théorie de l'information. La notion de d-séparation (Becker *et al.*, 1999), (Naïm *et al.*, 2007) joue un rôle important. En effet, le statut des noeuds d-séparés change par l'instanciation d'un élément de l'ensemble de noeuds. La quantité d'écoulement de l'information entre deux noeuds (I) peut être mesurée : soit en employant l'information mutuelle lorsque aucun noeud n'est instancié. Soit en employant l'information mutuelle conditionnelle, lorsque quelques autres noeuds sont instanciés.

Dans la figure 1 ci-dessous, nous montrons un modèle de connaissance extrait de nos données fixes, en utilisant l'algorithme K2 (Trabelsi *et al.*, 2010). Le modèle trouvé a pu détecter des relations entre les variables logiques comme la relation entre l'âge et l'antécédent, entre l'âge et cissue (le patient est décédé ou a survécu). Cependant le graphe obtenu, contient des liens " illogiques" entre les noeuds comme celui l'âge et la prise d'antibiotique). Nous avons aussi noté des liens manquants qui présentent des relations d'indépendance intéressantes (Par exemple, la relation entre result (infnos) et cissue).

Pour l'apprentissage des paramètres nous avons utilisé le calcul du maximum de vraisemblance. Nous avons construit les tableaux de probabilités conditionnelles qui présentent : $P(V_i | C)$, sachant que V_i est le noeud (sex, age1, periode_entr,...diag1), Les états de ces noeuds ont des valeurs discrètes, et C est la classe à prédire (cissue et infnos(result)). Dans la figure 2 nous allons présenter les probabilités $P(\text{age1}=\text{âgé} | \text{result}=\text{oui})$ (Trabelsi *et al.*, 2010).

3.3.2. Construction des modèles de connaissance des données temporelles (RBD)

Nous avons réalisé la première partie de notre système qui s'intéresse à l'apprentissage à partir des données fixes. Nous avons obtenu un Réseau Bayésien statique :

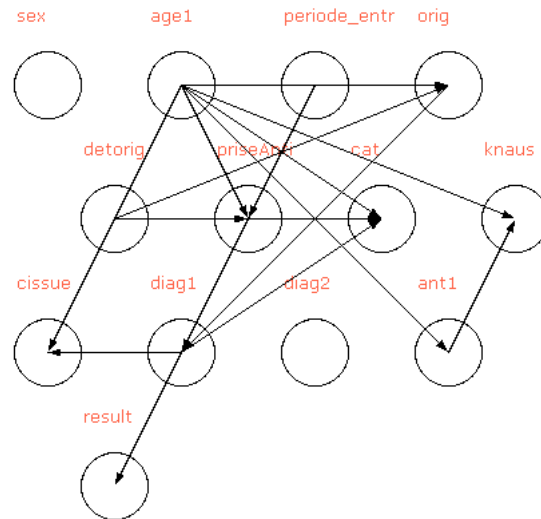


Figure 1. extrait par l'algorithme K2

age1/result	Non	Oui
adulte	0,560693641618497	0,428571428571429
agé	0,271676300578035	0,457142857142857
petit	0,167630057803468	0,114285714285714

Figure 2. La table de probabilité conditionnelle correspond au noeud age1

un graphe causal avec les tables de probabilités associées à chaque noeud. L'utilisation conjointe des probabilités et du graphe nous offre une famille de modèles qui ne sont pas très riches de connaissances utiles. Il faut donc passer à l'étape suivante qui consiste à apprendre les données temporelles et à élaborer des modèles de connaissances temporels qui sont plus riches que les modèles statiques. Nous allons présenter la structure du Réseau Bayésien dynamique obtenu par notre algorithme itératif et récursif d'apprentissage.

L'objectif de notre RBD est de prédire l'événement d'acquisition d'une infection nosocomiale par un patient et ceci tout au long de son hospitalisation, par le calcul de la probabilité quotidienne d'avoir un événement au futur sachant les observations (les actes et les examens infectieux effectués) dans le passé et au cours du jour courant. La figure figure 3-B montre qu'une transition du RB est composée par deux types d'arc : (1) Les arcs intra-tranche qui représentent l'interdépendance entre les variables d'une tranche. Nous avons utilisé pour cette structure celle du RB naïf puisque les variables

de chaque série de temps ($acte_1, \dots, acte_{10}, exinf_1, \dots, exinf_{30}$) n'ont aucune dépendance entre elles et qu'elles sont toutes connectées directement avec les deux noeuds cibles (de prédiction) qui sont result (ou infnos) et cissue. (2) les arcs d'inter-tranche qui représentent les arcs de temps. En effet, ces arcs symbolisent les interdépendances qui existent entre la variable elle-même mais pour des tranches de temps différentes et successives. Notre système a appris que le résultat de prédiction d'un jour t dépend directement du résultat de t-1. Aussi, nous avons inséré dans notre modèle temporel des connaissances données par l'expert en ce qui concerne les examens infectieux. Notre RBD est un RB qui modélise des distributions de la probabilité sur les collections semi-infinies de variables aléatoires (result, $act_i, \dots, exminfi_i, \dots, cissue$). Chaque modèle décrit les relations entre les différentes variables et le noeud cible "result_t". Le principe de notre RBD peut être défini comme un déroulement des transitions jusqu'à la fin de. Le résultat de la distribution des probabilités jointes est donné par :

$$P(result_{1:T}) = \prod_{t=1}^T \prod_{i=1}^N P(result_t^i | Pa(result_t^i)) \quad [3]$$

Où T est l'intervalle de temps d'hospitalisation et N est le nombre total des variables pour chaque tranche de temps. Nous obtenons un RBD final ayant comme structure celle représentée ci-dessous (Trabelsi *et al.*, 2010) :

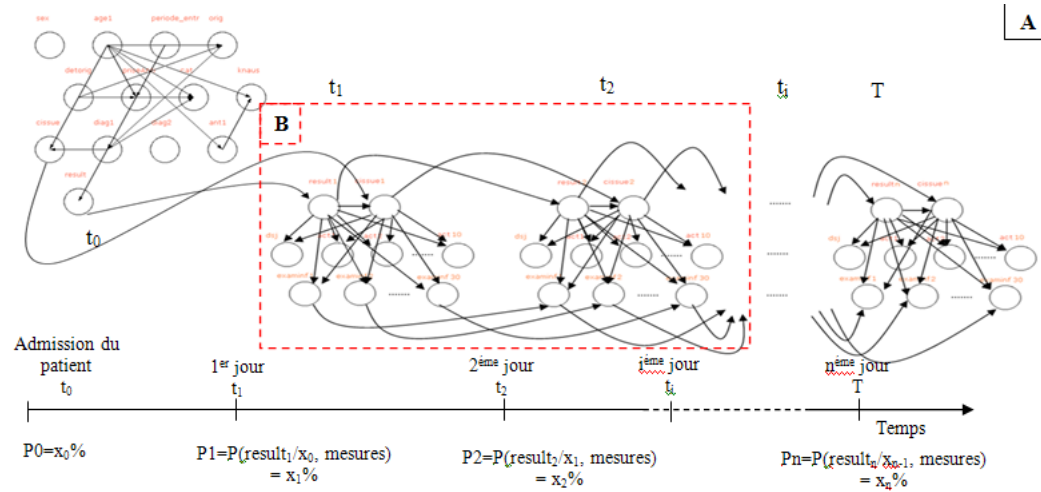


Figure 3. A :Le graphe causal de notre RBD. B : Représentation de deux tranches de temps de notre réseau dynamique

4. Résultats

4.1. Résultats de la prédiction dynamique

Notre étude a pour objectif de prédire la survenue d'IN chez un patient tout au long de sa hospitalisation. Les mesures quotidiennes enrichissent les modèles de prédiction. A chaque jour_t d'hospitalisation du patient nous avons pu prévoir son état au futur par une probabilité, qui sera utilisée, dans la prédiction du jour_{t+1}, avec ces observations mesurées.

La figure ci-dessous montre la prédiction effectuée sur un patient qui est resté 40 jours dans le service de réanimation. Pour le quel notre système a prédit la survenue d'une infection nosocomiale au 7^{ieme} jour de son hospitalisation et dans le 10^{ieme} jour le patient a effectivement attrapé une IN.

Sur la figure 4, nous avons présenté trois bulles :

id	dsj	CAT	PIV	Cv	exam_nF1	sens1	exam_nF2	sens2	exam_nF3	sens3	result
P216S1	1	1	1	1	0	0	0	0	0	0	0,0000000000000000
P216S2	2	1	0	1	0	0	0	0	0	0	0,1124085211940298
P216S3	3	1	0	1	0	0	0	0	0	0	0,0303951402081442
P216S4	4	1	0	1	0	0	0	0	0	0	0,2439185277152202
P216S5	5	1	0	1	0	0	0	0	0	0	0,4422679028990202
P216S6	6	1	0	1	0	0	0	0	0	0	0,1040919757550802
P216S7	7	1	0	1	0	0	0	0	0	0	0,6273599507500202
P216S8	8	1	0	1	0	0	0	0	0	0	0,1882410776020202
P216S9	9	1	0	1	0	0	0	0	0	0	0,0000000000000000
P216S10	10	1	0	1	1000	Résistant	134	Résistant	79	Résistant	out
P216S11	11	1	0	1	0	0	0	0	0	0	out
P216S12	12	0	0	1	0	0	0	0	0	0	out
P216S13	13	0	0	1	0	0	0	0	0	0	out
P216S14	14	0	0	1	1000	Sensible	134	Sensible	79	Résistant	out
P216S15	15	0	0	1	0	0	0	0	0	0	out
P216S16	16	0	0	1	0	0	0	0	0	0	out
P216S17	17	0	0	1	0	0	0	0	0	0	out
P216S18	18	0	0	1	0	0	0	0	0	0	out
P216S19	19	0	0	1	0	0	0	0	0	0	out
P216S20	20	0	0	1	0	0	0	0	0	0	out
P216S21	21	0	0	1	0	0	0	0	0	0	out
P216S22	22	0	0	1	0	0	0	0	0	0	out
P216S23	23	0	0	1	0	0	0	0	0	0	out
P216S24	24	0	0	1	0	0	0	0	0	0	out
P216S25	25	0	0	1	0	0	0	0	0	0	out
P216S26	26	0	0	1	0	0	0	0	0	0	out
P216S27	27	0	0	1	0	0	0	0	0	0	out
P216S28	28	0	0	1	0	0	0	0	0	0	out
P216S29	29	0	0	1	0	0	0	0	0	0	out
P216S30	30	0	0	1	0	0	0	0	0	0	out
P216S31	31	0	0	1	0	0	0	0	0	0	out
P216S32	32	0	0	1	0	0	0	0	0	0	out
P216S33	33	0	0	1	0	0	0	0	0	0	out
P216S34	34	0	0	1	0	0	0	0	0	0	out
P216S35	35	0	0	1	0	0	0	0	0	0	out
P216S36	36	0	0	1	1000	Sensible	134	Résistant	79	Résistant	out
P216S37	37	0	0	1	0	0	0	0	0	0	out
P216S38	38	0	0	1	0	0	0	0	0	0	out
P216S39	39	0	0	1	0	0	0	0	0	0	out
P216S40	40	0	0	1	0	0	0	0	0	0	out

Figure 4. Évolution de la prédiction d'attraper une infection nosocomiale pour un patient au cours du temps

- (A) indique les 40 identifiants des séries de temps relatives à la durée d'hospitalisation du patient en question ;
- (B) montre les différentes valeurs de probabilité prédite d'acquérir une IN et ceci pour 9 jours de son hospitalisation c-à-dire avant la détection d'une IN ;
- (C) après cette période le patient a acquis une IN. qui conforme au résultat du 7^{ieme} jour ;

La figure 4 présente les données temporelles concernant le patient de code P216. Notre base de données contient un identifiant pour chaque patient à chaque jour d'hospitalisation (ex :le code P216S14 correspond au 14^{ieme} jour d'hospitalisation du patient code P216). La variable dsj représente le nombre de jour d'hospitalisation, c'est une

Les résultats observés / Les résultats prédits	Oui	Non	Totale
Oui	5	2	7
Non	3	13	16
Totale	8	15	23

Tableau 1. *La matrice de confusion des résultats donnés par notre Réseau Bayésien dynamique*

variable très importante et qui a une action directe sur la prédiction. Plus la valeur de cette variable est grande plus la probabilité d'attraper IN est grande. La base contient aussi des données temporelles comme les actes et examens. il y a 10 types d'actes et 320 types d'examen infectieux. Chaque patient subit au maximum 10 actes et 30 examens par jour. Nous avons représenté dans la figure 4 quelques acronymes d'actes (CAT, PIV, CV), d'examen infectieux et le comportement du germe vis à vis de l'antibiotique (résistant ou sensible). Un autre champ contient le résultat de prédiction donné à chaque jour par notre système.

Cette prédiction est donnée sous forme d'une probabilité, plus cette probabilité augmente plus le danger d'attraper un IN est élevé. Nous remarquons que pour le patient P216, notre système a prédit une probabilité de 0,05 pour le premier jour d'hospitalisation, cette probabilité a augmenté à 0,11 puis à 0,63 pour respectivement le 2^{ème} et le 3^{ème} jour. A ce stade le médecin peut prédire que ce patient attrapera une IN. Cette probabilité démunie dans les jours suivants jusqu'à atteindre une probabilité de 0,10. Au 7^{ème} et 8^{ème} jour cette dernière a augmenté de nouveau pour dépasser 0,6. En ce moment, le médecin est sûr que ce patient va attraper une IN dans les jours suivants. Ce qui est réellement passé, en effet ce patient a bien attrapé une IN le 10^{ème} jour de son hospitalisation. Nous avons classé ce patient comme un patient infecté en lui assigné la valeur "oui" dans la base de résultats.

4.2. Les résultats de la base de test

Dans cette section, nous rappelons que notre base contient 280 cas. Nous n'avons tenu compte pour l'évaluation de notre système que des patients qui ont un séjour plus de 48 heures à l'hôpital (une infection est considérée comme nosocomiale si elle a apparue après 48h d'hospitalisation). Nous avons utilisé des cas de patients pour le test différents des cas utilisés dans l'apprentissage (200 cas pour l'apprentissage et 23 cas pour le test).

Le tableau 1 illustre clairement que le taux des reconnaissances positives correctes est élevé, 5 patient sont correctement classés parmi 8 qui ont attrapé une IN. Nous avons, aussi, un taux des reconnaissances négatives correctes grand, 2 cas sont incorrectement classés parmi 15 patients non infectés (voir Tableau 1). Nous montrons les résultats obtenus de la prédiction dynamique dans la matrice de confusion (Tableau 1) : Nous avons calculé les taux d'évaluation à partir des résultats de prédiction obtenus par

notre structure élaborée par le RBD, nous avons trouvé que le taux de classification était correcte à 0.78, le pouvoir de prédiction positif = 0.62, Le pouvoir de prédiction négatif = 0.86, Sensibilité = 0.71, Spécificité = 0.8125 et les chances proportionnelles = 3,6.

5. Conclusion

Dans ce travail, nous avons proposé une approche d'extraction des connaissances à partir des données temporelles. Notre approche consiste à appliquer les RB classiques sur les données fixes et les RBD sur les données temporelles. Nous avons validé notre approche sur un cas réel. Nous avons ainsi développé un système de prédiction quotidienne de l'IN chez les patients dans le service de réanimation de l'hôpital Habib Bourguiba de Sfax, Tunisie. Nous avons pu extraire des modèles de connaissance et de les transformer automatiquement comme des résultats probabilistes, quantitatives et qualitatives pour la prédiction.

Comme perspectives nous envisageons de mettre en place ce système dans d'autre service, adopter à d'autre types de prédiction. Et enfin améliorer notre approche en ce qui concerne l'apprentissage dynamique et temporelle des modèles de connaissances.

6. Bibliographie

- Antonucci A., Zaffalon M., « Decision-theoretic specification of credal networks : A unified language for uncertain modeling with sets of Bayesian networks », *International Journal of Approximate Reasoning*, vol. 49, p. 345-361, 2008.
- Becker A., Naïm P., *les Réseaux Bayésiens : Modèles graphiques de connaissance*, Editions Eyrolles, Paris, 1999.
- Brossette S., Sprague A., Jones W., Moser S., « A data mining system for infection control surveillance », *Methods Inf Med*, vol. 39, n° 5, p. 303-310, 2000.
- Correa M., Bielza, Pamies-Teixeira J., « Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process », *Expert Systems with Applications*, vol. 36, n° 3, p. 7270-7279, 2009.
- Courtney J., « Decision making and knowledge management in inquiring organization : toward a new decision-making paradigm for DSS », *Decision Support Systems*, vol. 31, p. 17-38, 2001.
- Fayyad U., « Automating the Analysis and Cataloging of Sky Surveys », *Advances in Knowledge Discovery and Data Mining*, NRC, Valencia, Spain, p. 471-494, 1996.
- Garner J., Jarvis W., Emori T., Hogan T., Hugues J., « CDC definitions for nosocomial infections », *Am. journal of Infection Control.*, vol. 16, n° 3, p. 128-140, 1988.
- Jordan M., « Graphical Models », *Statistical Science (Special Issue on Bayesian Statistics)*, vol. 19, p. 140-155, 2004.
- Kallel H., Bouaziz M., Ksibi H., Chelly H., Hmida C., Chaari A., Rekik N., Bouaziz M., « Prevalence of hospital-acquired infection in a Tunisian Hospital », *Journal of Hospital Infection*, vol. 59, p. 343-347, 2005.

- Keen P., Scott M., « An organizational perspective », *Decision Support, Systems Addison-Wesley Publishing Company*, 1978.
- Lefébure R., Venturini, *Data Mining : Gestion de la relation client, Personnalisation des sites Web*, Eyrolles, Paris, 2001.
- Lepreux S., Approche de Développement centré décideur et à l'aide de patrons de Systèmes Interactifs d'Aide à la Décision, Thèse de doctorat, l'Université de Valenciennes et du Hainaut-Cambrésis, 2005.
- Leray P., Réseaux Bayésiens : apprentissage et modélisation de systèmes complexes, habilitation à diriger les recherches, Université de Rouen, 2006.
- Murphy K., Dynamic Bayesian Networks : Representation, Inference and Learning, Thèse de doctorat, University of California, Berkeley, 2002.
- Naïm P., Wuillemin P., Leray P., Pourret O., Becker A., *Les Réseaux bayésiens*, Editions Eyrolles, Paris, 2007.
- Neapolitan R., *Learning Bayesian Networks*, Pearson Education, upper saddle River, 2004.
- Pearl J., *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Mateo, 1988.
- Peelen L., Keizer N. D., Jonge E. D., Bosman R., Abu-Hanna A., Peek N., « Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in the Intensive Care Unit », *journal of Biomedical Informatics*, vol. 43, n° 2, p. 273-286, 2010.
- Pernkopf F., « Bayesian network classifiers versus selective K-NN classifier », *Pattern Recognition*, vol. 38, p. 1-10, 2005.
- Scott M., « Computer based support for decision making », *Management decision systems*, Harvard University, Boston, MA, USA, 1971.
- Shahar Y., « Dimensions of time in illness : an objective view », *Ann Intern Med*, vol. 132, p. 45-53, 2000.
- Sprague R., Carlson E., « Building Effective Decision Support Systems », *Prentice-Hall*, Harvard University, Boston, MA, USA, Inc, Englewood Cliffs, 1982.
- Trabelsi G., Ayed M. B., Alimi A., « Système d'extraction des connaissances à partir des données temporelles basé sur les Réseaux Bayésiens Dynamiques », *RNTI E19 Extraction et Gestion des Connaissances EGC*, cépadués edition, Hammamet, Tunisie, p. 241-246, JAN, 2010.