

I modelli grafici esplorativi per l'analisi del fabbisogno di assistenza ospedaliera informale

Mario Bolzan, Adriana Brogini, Debora Slanzi¹

Dipartimento di Scienze Statistiche, Università di Padova.

Riassunto. La funzione prevalente dell'attività di ricerca applicata nelle scienze sociali e sanitarie è quella di cercare di individuare le determinanti che definiscono il sistema e le loro relazioni. L'approccio fornito dalla *Knowledge Discovery* offre uno dei più promettenti in tale contesto. Il *Knowledge Discovery in Database* (KDD) indica l'intero processo di ricerca di nuova conoscenza formale procedendo da un database del fenomeno in studio. Esistono molti formalismi di rappresentazione della conoscenza estratta: in questo lavoro l'attenzione è rivolta alle reti Bayesiane, vale a dire ai modelli grafici di dipendenza probabilistica. L'analisi esplorativa svolta nel lavoro intende evidenziare la complessità dell'assistenza informale intra- ed extra-ospedaliera espressa dai ricoverati, sottolineando il forte contributo che alcune variabili svolgono nella costruzione delle relazioni che si attivano all'interno di questo sistema.

Parole chiave. *Knowledge Discovery in Database*; Rete Bayesiana; Apprendimento del modello; Assistenza ospedaliera informale.

1. Introduzione

La ricerca applicata nelle scienze socio-sanitarie si configura spesso in termini complessi e multidimensionali. La funzione prevalente della ricerca applicata in queste aree è quella di cercare di comprendere le determinanti che definiscono il sistema in misura ottimale e le relazioni che fra queste si attivano. L'approccio della *Knowledge Discovery* (Hastie *et al.*, 2001; Mitchell, 1997; Fayyad *et al.*, 1996) è uno dei più informativi e interessanti a questo proposito.

¹ Il presente lavoro è stato finanziato nell'ambito del progetto "Transizioni Università-Lavoro e valorizzazione delle competenze professionali dei laureati: modelli e metodi di analisi multidimensionali delle determinanti", cofinanziato dal MIUR. Coordinatore nazionale e del Gruppo di Padova è L. Fabbris. Il lavoro è opera comune degli autori.

L'analisi da attivare si configura come un processo di identificazione degli aspetti fondamentali (variabili) del sistema e delle interazioni che informano per il problema in esame. La costruzione del modello richiede il coinvolgimento di esperti, oltre alla disponibilità di una base dati da elaborare mediante opportuni algoritmi computazionali e una modalità di rappresentazione dei risultati che ne faciliti l'interpretazione.

L'applicazione dell'approccio KDD, che si presenta dettagliatamente nel Par. 2, procede dalla selezione e trasformazione dei dati e si sviluppa durante l'applicazione di funzioni e algoritmi di *data mining* fino alla specificazione di un modello in grado di descrivere, analizzare e prevedere il fenomeno oggetto di studio. Il modello che rappresenta le dipendenze significative tra le variabili del dominio è specificato in termini strutturali come rappresentazione grafica della dipendenza locale delle variabili e in termini quantitativi rispetto alla forza della dipendenza.

Fra le diverse formalizzazioni di rappresentazione della conoscenza prodotta, l'attenzione è rivolta alle reti Bayesiane che rappresentano modelli grafici di dipendenza probabilistica, utili alla descrizione delle conoscenze prodotte.

Il presente lavoro si pone l'obiettivo di svolgere un'analisi esplorativa sul fabbisogno di assistenza informale del paziente durante il ricovero ospedaliero utilizzando le reti Bayesiane. Una rete è un mezzo per la comprensione delle relazioni che esistono tra le variabili del dominio.

Il lavoro è organizzato nel seguente modo: nel Par. 2 si introducono le reti Bayesiane, nel Par. 3 viene descritto il problema oggetto di studio e il database usato nell'analisi, nel Par. 4 si presentano i risultati e si discute il contributo informativo di queste tecniche per spiegare il fenomeno complesso oggetto di studio.

2. Le reti Bayesiane

Una rete Bayesiana è un modello grafico che rappresenta una distribuzione di probabilità congiunta $P(\mathbf{X})$ per un insieme finito non vuoto di variabili aleatorie discrete $\mathbf{X}=\{X_1, \dots, X_n\}$ tale che ognuna di esse rappresenta una variabile del dominio oggetto di studio². Una rete rappresenta $P(\mathbf{X})$ attraverso la fattorizzazione

$$P(\mathbf{X}) = \prod_{X \in \mathbf{X}} P(X | Pa(X)).$$

² Formalmente, una rete Bayesiana per \mathbf{X} è una coppia $B=(G, \theta)$, dove G è un grafo diretto aciclico i cui nodi corrispondono a variabili casuali in \mathbf{X} , e θ sono i parametri che specificano una distribuzione di probabilità condizionata per ogni nodo $X \in \mathbf{X}$ dati i suoi parenti $Pa(X)$ in G , $P(X|Pa(X))$. Dato un grafo G e due nodi X e Y del grafo, si dice che Y è un parente di X se esiste una freccia diretta da Y a X e si indica $Y=Pa(X)$.

Il grafo diretto aciclico (DAG) G della rete permette di fare considerazioni in merito alla struttura delle (in)dipendenze condizionate³ tra le variabili casuali in \mathbf{X} . Queste (in)dipendenze condizionate possono essere lette da G attraverso il criterio di *d-separazione* (Lauritzen, 1996).

La proprietà basilare di una rete Bayesiana è che ogni variabile è condizionatamente indipendente dai suoi discendenti, dove un discendente di un nodo è definito o come un figlio del nodo, oppure un discendente di uno dei suoi figli, dati i suoi parenti (Pearl, 1988). Questa proprietà, detta locale di Markov, esprime un insieme minimo di relazioni di indipendenza che esistono tra ogni nodo e i suoi non-discendenti, fissata una rete (Fig. 1).

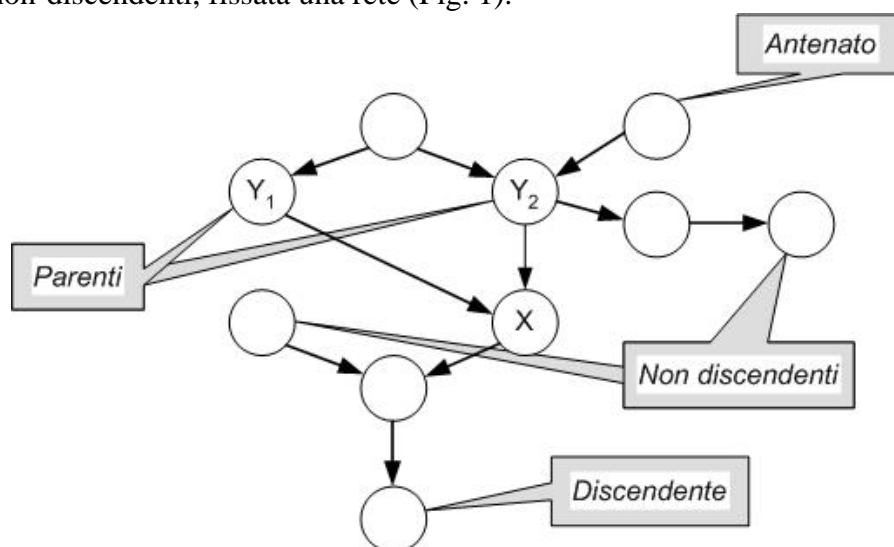


Figura 1: Esempio di rete Bayesiana, in cui per il nodo X vengono specificati l'insieme dei parenti, degli antenati, dei discendenti e non discendenti di X .

Una rete Bayesiana è dunque un modello grafico che rappresenta in forma compatta le relazioni probabilistiche tra un insieme di variabili, utile per codificare l'incertezza della conoscenza nei sistemi esperti (Heckerman *et al.*, 1995).

Le reti Bayesiane presentano una consistenza teorica per processi incerti, in quanto la teoria probabilistica su cui sono basate permette il calcolo dell'inferenza in modo definito, risultano robuste, nel senso che piccole alterazioni del modello non influiscono sul comportamento dell'intero sistema, sono flessibili nelle applicazioni, in quanto possono essere usate per risolvere problemi di classificazione, di configurazione, previsione o di *data mining*. Quest'ultimo risulta essere il principale

³ Una freccia tra due nodi indica la relazione di dipendenza diretta condizionata che esiste tra essi. Analogamente la mancanza di una freccia tra due variabili ne indica la indipendenza condizionata.

obiettivo di questa analisi, che esamina le distribuzioni condizionate, le dipendenze e le correlazioni trovate nella modellazione del processo.

Le reti e le tecniche di statistica Bayesiane facilitano la combinazione della conoscenza a priori del dominio oggetto di studio con i dati. Nella modellizzazione Bayesiana, infatti, la conoscenza dell'esperto può essere incorporata come distribuzione a priori⁴. Tutti i parametri della rete hanno una chiara interpretazione semantica, per cui il modello risultante può essere analizzato e spiegato in termini comprensibili agli utilizzatori.

3. Il problema e il database utilizzato nell'analisi

Nel nostro Paese la normativa assegna all'assistenza ospedaliera l'obiettivo di "garantire a tutti i soggetti assistiti dal Servizio Sanitario Nazionale l'accesso ai ricoveri ospedalieri per trattare condizioni patologiche indifferibili che necessitano di interventi diagnostico-terapeutici di emergenza e di urgenza, patologie acute non gestibili in ambito ambulatoriale e/o domiciliare, nonché condizioni patologiche di lunga durata, che richiedono un trattamento diagnostico-terapeutico non erogabile in forma extraospedaliera" (DPCM 29 novembre 2001, GU n.33, 8 febbraio 2002).

La programmazione delle risorse, per l'assistenza ospedaliera, mira a riqualificare le politiche assistenziali attraverso una riorganizzazione e un riequilibrio degli investimenti e delle spese. Inoltre si pone l'obiettivo di potenziare e qualificare la rete informale di assistenza extraospedaliera a domicilio, in famiglia e nelle residenze sanitarie assistite. Tutto ciò ha prodotto cambiamenti nell'organizzazione e nella gestione dei servizi, cosa che si può notare anche dalla lettura di indicatori del processo del sistema dei servizi ospedalieri in Italia negli ultimi anni (ISTAT, 2001).

L'assistenza al familiare ricoverato, in particolare se malato terminale, rappresenta un impegno non trascurabile per la famiglia (Litman, 1974). In Italia, almeno un terzo dei ricoverati ha bisogno di assistenza informale (familiari, volontari o assistenti retribuiti) durante il ricovero ospedaliero ed è ragionevole supporre che tale servizio non si concluda con le dimissioni formali dall'istituto di cura. Esiste una consistente associazione fra fabbisogno di assistenza personalizzata e informale di cui il paziente usufruisce e la durata del ricovero, espressione della complessità delle ragioni e modalità del ricovero, ma anche di quanto il sistema ospedaliero abbia bisogno di risorse esterne (reti informali) nell'assistenza dei pazienti (Bolzan, 1989).

In questo lavoro, con "fabbisogno" s'intende sia la domanda espressa dal paziente che l'utilizzo dei servizi formali e informali disponibili e offerti per

⁴ Per approfondimenti teorici sulle reti Bayesiane si veda Heckerman (1996), Jensen (1996), Buntine (1996b) e Neapolitan (2004).

l'assistenza ospedaliera. In Italia, il problema dell'assistenza assume ulteriore consistenza e interesse se si considera che il 25,8% delle associazioni di volontariato presenti nel territorio nazionale si occupa di assistenza ai malati, in maniera fortemente diversificata a livello territoriale (31,6% al Centro e circa il 24% altrove) e che su tutte queste organizzazioni ben il 18,5% delle prestazioni è di natura sanitaria (con punte del 24,7% al Centro e il 13,8% al Nord Est). Inoltre, fra le donne lavoratrici in coppia (cioè con il coniuge lavoratore) ben il 54% deve sostenere attività lavorative superiori alle 60 ore (comprensive del lavoro domestico ed extra imputabile agli impegni familiari) contro il 15% fra gli uomini. Questi elementi illustrano una parte del contesto all'interno del quale la famiglia italiana si trova a dover operare nelle condizioni di necessità sostenute in occasione del ricovero ospedaliero di un familiare (Pless, 1984).

Si può quindi comprendere l'importanza dell'analisi dei fattori che determinano l'utilizzo dell'assistenza informale, anche durante il ricovero. L'analisi può risultare molto utile nei processi di valutazione delle diverse strategie di cura e assistenza post ricovero da attivarsi.

La base dati utilizzata nell'analisi è stata selezionata dall'Indagine campionaria Multiscopo sulle famiglie fatta dall'ISTAT nel 1999-2000 denominata *Sulle condizioni di salute e ricorso ai servizi sanitari della popolazione italiana*. Questa indagine ha avuto lo scopo di cogliere il modo in cui vive la popolazione italiana, qual è il suo stato di salute e il ricorso ai servizi sanitari (www.istat.it)⁵.

Per realizzare gli obiettivi conoscitivi prefissati sul fabbisogno di assistenza informale durante il ricovero ospedaliero, si è dapprima proceduto ad una adeguata selezione ed elaborazione del database a disposizione. Poiché la rilevazione è stata svolta sulle famiglie e sui singoli componenti, nella costruzione del database utilizzato nell'analisi sono state selezionate solo le famiglie con almeno un membro che avesse avuto un periodo di ricovero ospedaliero durante l'anno di rilevazione; per le famiglie in cui si dichiarava la presenza di più di una persona ricoverata, si è scelto in modo casuale una fra queste.

Far parte di una famiglia con più di un componente ricoverato in ospedale potrebbe essere una variabile significativa ai fini dell'esplorazione del fenomeno in esame. Questa informazione è stata tuttavia trascurata in quanto nel database si è riscontrata una bassa frequenza di famiglie in questa condizione.

Il database così formato è costituito da 5056 records, che corrispondono al numero di soggetti ricoverati in ospedale appartenenti ad altrettante famiglie. Si considerano le osservazioni su 21 variabili inerenti al problema dell'assistenza

⁵ Il campione di famiglie è casuale a due stadi con stratificazione delle unità di primo stadio (Comuni). L'indagine ha raggiunto 21.931 famiglie per un totale di 62.461 individui. Le informazioni sono state raccolte con intervista diretta o tramite autocompilazione. L'unità di rilevazione è costituita dalla famiglia di fatto associata alla famiglia anagrafica campionata.

informale. Alcune di queste sono state costruite attraverso una riclassificazione delle variabili originarie, rilevate nell'indagine nazionale. Il database ottenuto non presenta dati mancanti, le variabili sono discrete, o sono state opportunamente discretizzate, ognuna con un numero finito di stati (modalità).

Si presentano le variabili utilizzate nell'analisi e i possibili stati assunti. Le variabili selezionate sono divise in tre sottogruppi, in accordo alla loro funzione nel dominio studiato.

- Variabili di contesto: sono le variabili generali che esprimono le caratteristiche socio-demografiche dell'unità statistica.
 - Età: 0-14, 15-29, 30-44, 45-59, 60-75, 75+
 - Sesso: Maschio, Femmina
 - Ripartizione geografica :Nord-Est, Nord-Ovest, Centro, Sud, Isole
 - Professione: (riferita al capofamiglia) Occupato, Pensionato, Altro
 - Titolo di studio capofamiglia: Nullo, Elementari, Media, Diploma, Laurea
 - Stato civile: Celibe / Nubile, Coniugato, Altro
 - Parentela: Capofamiglia, Altro
 - Numero di componenti della famiglia: 1, 2, 3, 4+
 - Tipologia familiare: Famiglia senza nucleo, Monogenitore, Coppia senza figli, Coppia con figli, Famiglie con due nuclei
 - Rilevazione: Settembre 1999, Dicembre 1999, Marzo 2000, Giugno 2000
 - Malattie croniche: Mai sofferto, Almeno una in passato, Sì, almeno una
 - Invalidità: No, Sì
- Variabili intermedie: sono variabili concomitanti che interagiscono con le variabili contesto per condizionare la variabile target.
 - Interventi: No, Sì
 - Modalità di ricovero: Prenotazione, Tramite pronto soccorso, Tramite ospedale, Altro
 - Giorni intercorsi tra prenotazione e ricovero: 0-7, 8-15, 16-30, 31-60, 60+
 - Motivi del ricovero: Accertamenti, Malattie / Intervento, Incidente, Altro
 - Numero di notti in ospedale: 0-4, 5-7, 8-14, 15+
 - Struttura: Pubblica, Privata
 - In comune: Sì, Altro comune della provincia, Altra provincia della regione, Fuori regione. Indica se la struttura dove è avvenuto il ricovero si trova o meno nel comune di residenza del paziente.
 - Provveduto al pagamento di un contributo da versare all'ospedale per il ricovero effettuato: No, Sì tramite assicurazione
- Variabile target: Assistenza informale. Definita e rilevata dall'ISTAT attraverso la variabile "Assistenza da parte di persone esterne i servizi ospedalieri durante le notti di ricovero" con modalità "Sì" e "No". Nel questionario originale si individuano come possibili persone da cui ricevere assistenza:

- familiari o altre persone conviventi;
- parenti non conviventi;
- amici, colleghi o vicini di casa;
- persone non retribuite appartenenti a gruppi di volontariato;
- personale sanitario a pagamento;
- altro personale a pagamento.

4 Apprendimento della rete Bayesiana

Apprendere una rete Bayesiana significa individuare (Spiegelhalter e Lauritzen, 1990; Buntine, 1994; Heckerman *et al.*, 1995):

- i parametri θ delle distribuzioni di probabilità condizionata;
- il modello grafico di dipendenze: la struttura del DAG G .

Quando il grafo non è noto, si ricerca una stima della topologia, intesa in questo contesto come i legami della rete. Apprendere la struttura di una rete dai dati è noto come problema di *model selection* nel senso che ogni rete corrisponde ad un modello distinto e si deve selezionare, in base ai dati, una tra le reti in esame. Per l'apprendimento della struttura della rete Bayesiana di è adottato l'approccio *search and score*: si seleziona un "buon" modello fra tutti quelli possibili e lo si usa come fosse quello corretto. La "bontà" del modello viene misurata in accordo con un criterio: specificatamente, viene misurato il grado con cui una struttura si adatta alla conoscenza a priori e ai dati.

Si definisce dunque una funzione *score* che ha il compito di codificare la bontà del modello su cui viene calcolata. Nell'articolo si adotta la funzione *score Bayesiana* discussa per la prima volta in Cooper e Herskovitz (1992) e sviluppata in Heckerman *et al.* (1995) e Buntine (1996a): $Score(B, D) = P(B/D)$. Essa rappresenta la probabilità a posteriori di B dato il database D (Buntine, 1996b).

Successivamente, attraverso un algoritmo di ricerca, si trova quale fra i modelli candidati risulta essere il migliore, ovvero riceve *score* maggiore. In letteratura esistono diversi algoritmi per la fase di ricerca, all'interno dello spazio delle strutture possibili, come l'algoritmo *greedy K2* (Cooper e Herskovitz, 1992), un'estensione dello stesso proposta da Singh e Valtorta (1995), l'algoritmo *hill-climbing* e altre varianti della ricerca *greedy* proposte e valutate da Heckerman *et al.* (1995).

Per il problema oggetto di studio, e per lo scopo di questa analisi, la struttura della rete è ignota, quindi il processo è una combinazione di apprendimento dei parametri (quantitativo) e di struttura (qualitativo).

Il software utilizzato nell'analisi, *Bayesware Discoverer*, fornisce come funzione *score* la log-verosimiglianza marginale⁶ e utilizza l'algoritmo *greedy K2* come metodo di ricerca (Bayesware Limited, 2000).

L'algoritmo *greedy K2* richiede un ordinamento delle variabili⁷. Si cerca, dunque, l'ordinamento che porta all'apprendimento della rete con *score* più alta. Per ogni sottogruppo di variabili descritto nel Par. 3 sono costruite tutte le possibili reti ottenute con la permutazione delle variabili nel sottogruppo. Si fissa l'ordinamento delle variabili presenti in ogni sottogruppo uguale a quello a cui corrisponde la rete che presenta *score* maggiore fra tutte quelle ottenute permutando l'ordine delle variabili.

La rete finale è prodotta introducendo di volta in volta le variabili in base all'ordinamento precedentemente trovato nei sottogruppi: prima le variabili di contesto, poi quelle intermedie e per ultima quella di target. Il migliore ordinamento trovato è quello già presentato nel Par. 3. Ciò presuppone che variabili che seguono nell'ordine non possono essere considerate come parenti di quelle che le precedono. La rete Bayesiana⁸ risultante è rappresentata nella Fig. 2.

Una conclusione evidente dalla Fig. 2, è il ruolo centrale assunto dall'età del ricoverato. Questa variabile interagisce con le componenti sociali della famiglia come la professione (quindi il titolo di studio) del capofamiglia, con la composizione della famiglia stessa ed infine con la presenza di malattie croniche e altre forme di invalidità nel paziente che possono comportare riduzione o perdita di autonomia.

Come si può notare, esistono numerose *relazioni dirette* (freccie) tra le variabili che sottolineano come il problema sotto studio sia effettivamente espressione di un sistema sottostante complesso. Oltre alle relazioni dirette, si evidenziano anche molte *relazioni indirette* che si verificano quando si passa da un nodo ad un altro non direttamente attraverso la freccia che li collega, ma seguendo un cammino, inteso come un insieme di frecce che coinvolgono altre variabili. Nella

⁶ Assumendo che non si abbia nessuna informazioni a priori sulla topologia della rete, il che equivale a imporre una distribuzione a priori di tipo uniforme per la struttura, risulta che la probabilità a posteriori della struttura dato il database è proporzionale alla verosimiglianza marginale. Senza perdita di generalità può essere dunque utilizzata la log-verosimiglianza marginale come *score* del modello.

⁷ L'ordinamento delle variabili richiesto dall'algoritmo permette di ridurre lo spazio di ricerca delle possibili strutture che possono essere costruite utilizzando le variabili specificate nel dominio.

⁸ Per facilitare la lettura della rete si ricorda che: ogni nodo corrisponde ad una variabile del dominio oggetto di studio, ogni freccia fra i nodi descrive l'influenza diretta fra le variabili interessate. Una freccia uscente da una variabile e che punta verso un'altra, rappresenta la diretta dipendenza stocastica della prima (parente) sulla seconda (figlio). Ogni nodo ha associata una distribuzione di probabilità condizionata, rappresentata da una tabella (CPT), che indica le probabilità assunte da ogni nodo figlio in ognuno dei suoi differenti stati, per ogni combinazione di stati dei suoi parenti (parametri della rete). Se la variabile non ha parenti, la CPT associata è una tabella di probabilità marginali.[0]

Fig. 2 si può notare come esista una relazione tra la variabile Età e la variabile Interventi o attraverso la variabile Invalidità o attraverso la variabile Stato civile.

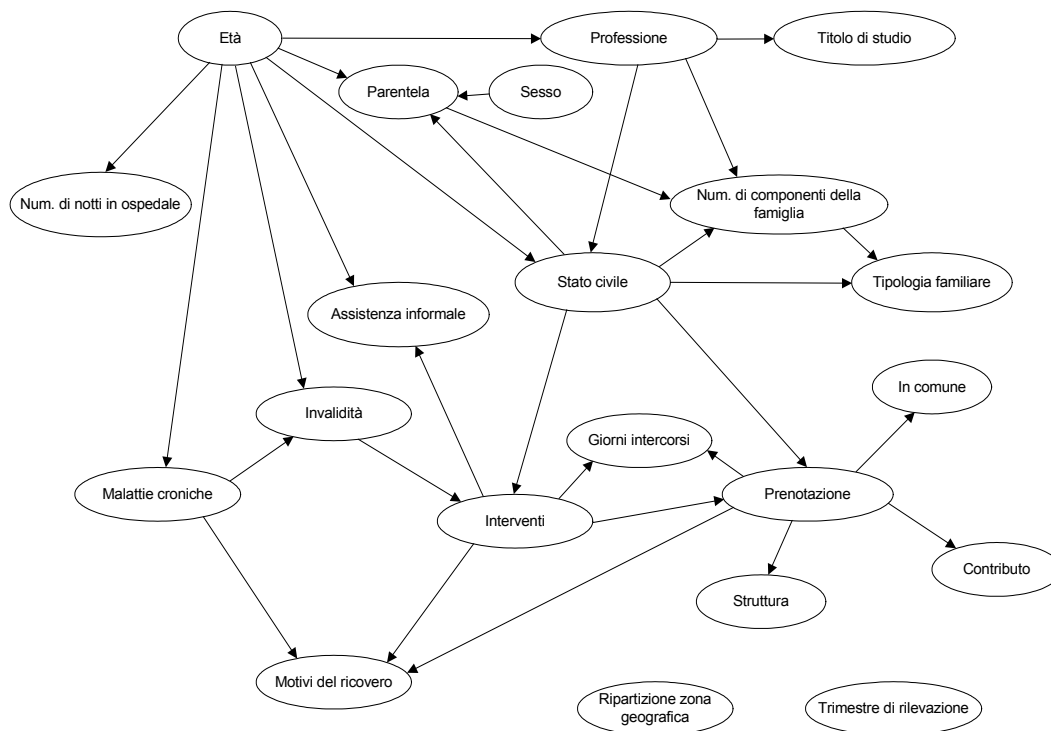


Figura 2. La rete Bayesiana

È, ovviamente, più significativa la relazione di (in) dipendenza che si legge attraverso una freccia diretta che collega due o più variabili, ma è altrettanto significativo tenere in considerazione anche le relazioni indirette. In questo lavoro l'attenzione è limitata alla lettura e all'interpretazione di alcune fra le relazioni evidenziate dalla rete. In particolare per l'obiettivo prefissato, la considerazione è stata limitata alle relazioni dirette che si attivano con la variabile Assistenza informale e alcune delle più significative relazioni che si sviluppano attorno alla stessa o alle variabili direttamente collegate ad essa. Si considera dunque la relazione diretta che la variabile Assistenza informale ha con l'Età e con Interventi, e per l'influenza di quest'ultima sulla variabile oggetto di studio, ci si focalizza anche sull'analisi delle relazioni che essa attiva con le variabili Invalidità e Stato civile.

Nel seguito vengono descritti i risultati delle elaborazioni mediante misure di probabilità⁹ e di rischio relativo (Armitage e Berry, 1996) utili per comprendere la differenziale forza esplicativa delle variabili o modalità coinvolte nell'analisi.

Si osserva come lo stato civile (Tab. 1), utile nell'indicare il microcosmo di riferimento del paziente, evidenzia particolari associazioni con l'invalidità e l'intervento (variabili queste che nella rete sono altamente associate al fabbisogno di assistenza durante il ricovero). Trattandosi di misure di probabilità, questi rapporti possono essere interpretati come misure del rischio relativo.

Per lo sposato/a (ragionevolmente adulto o anziano) la probabilità di aver avuto bisogno di un intervento durante l'anno di riferimento è per un non invalido 1,315 volte (*Rischio Relativo 1* nella Tab. 1) quella di un invalido; questa relazione appare ragionevole se si considera che l'invalido, per propria condizione, riceve prevalentemente un'assistenza continua, sistematica, programmata e generalmente riabilitativa, così che l'intervento rappresenta generalmente una soluzione radicale, spesso urgente, e scelta quando non vi siano alternative immediate.

La misura di questo rischio sale a 1,580 fra i single non invalidi rispetto agli invalidi (in media più giovani degli sposati) dove l'intervento è comunque meno frequente e, a parità di altri fattori, è ragionevolmente indicativo di una condizione di maggior gravità. La probabilità di subire un intervento è sempre minore (*Rischio Relativo 2* nella Tab. 1) se il paziente non è sposato (ad esempio bambino o comunque ragionevolmente non anziano). Queste evidenze possono aiutare a comprendere quanto emerso rispetto al fabbisogno di assistenza: la dimensione "familiare" svolge dunque un ruolo diverso rispetto a quello dei single.

Tabella 1. Distribuzione delle probabilità di intervento e dei rischi relativi condizionati a Invalidità = NO, per stato civile

| Stato civile | Interventi=SI (*) | <i>Rischio Relativo 1</i> | <i>Rischio Relativo 2</i> |
|--------------|-------------------|---------------------------|---------------------------|
| Sposato | 0,442 | 1,315 | 1,000 |
| Single | 0,384 | 1,580 | 0,869 |
| Altro | 0,326 | 1,482 | 0,737 |

(*) Probabilità di Interventi =SI condizionata a Invalidità =NO e allo Stato Civile

Rischio Relativo 1 = Rischio relativo di avere Intervento in base all'Invalidità (No vs Si);

Rischio Relativo 2 = Rischio relativo di avere Intervento in base allo Stato Civile

Lo Stato civile, espressione della relazione e del ruolo che il soggetto ha nella famiglia (moglie/marito), si prefigura come potenziale predittore delle modalità del sistema di assistenza informale che si può attivare durante il ricovero ed è ugualmente importante anche nelle fasi organizzative precedenti, quali la modalità di prenotazione ospedaliera.

⁹ Le probabilità di cui si fa riferimento rappresentano i parametri appresi della rete.

È comunque ancora l'intervento subito durante il ricovero, combinato con l'età, che esprime con particolare intensità il fabbisogno di assistenza durante il ricovero (Tab. 2). Si noti come la probabilità di aver bisogno di assistenza è elevata e maggiore fra coloro che hanno subito un intervento (con prevalenze consistenti fra i degenti in età pediatrica o adolescenziale).

Al crescere dell'età, la probabilità di fabbisogno (fra coloro che hanno subito intervento) diminuisce fino a quasi 60 anni per ricrescere senza comunque raggiungere i livelli dell'età pediatrica. Fra costoro infatti, il rischio relativo (Tab. 2) di averne bisogno è da circa 1,5 a 3 volte superiore rispetto a coloro che non hanno avuto l'intervento (massimamente nell'età 30-44 anni dove l'intervento durante il ricovero è giustificato da condizioni di ragionevole serietà). Considerando come dimensione standard la probabilità a 30-44 anni di aver bisogno di assistenza a condizione di aver subito un intervento, si osserva che alle altre età il rischio è sempre maggiore e massimo in tenera età (*Rischio Relativo 2* in Tab. 2). Infine emerge da successive analisi che il rischio di aver bisogno di assistenza notturna durante un ricovero ospedaliero - con intervento - in un bambino, è circa 9 volte quella di un ricoverato adulto di 30-44 anni che non ha subito l'intervento.

Tabella 2. Distribuzioni di probabilità condizionate e dei rischi relativi condizionati a Intervento=SI, per età

| Età | Assistenza Informale=SI (*) | Rischio Relativo 1 | Rischio Relativo 2 |
|-------|-----------------------------|--------------------|--------------------|
| 0-14 | 0,898 | 1,592 | 2,887 |
| 15-29 | 0,361 | 1,769 | 1,161 |
| 30-44 | 0,311 | 3,079 | 1,000 |
| 45-59 | 0,336 | 2,400 | 1,080 |
| 60-74 | 0,420 | 2,282 | 1,350 |
| 75+ | 0,529 | 1,396 | 1,701 |

(*) Probabilità di assistenza informale =SI condizionata a Interventi =SI e all'Età

Rischio Relativo 1 = Rischio relativo di avere assistenza informale in base all'Intervento(Si vs No)

Rischio Relativo 2 = Rischio relativo di avere assistenza informale in base all'Età

Simmetricamente, una degenza senza intervento vede associata una probabilità di autonomia da parte del ricoverato, cioè di non aver bisogno di assistenza, che solo nella prime classi d'età supera anche di 4 volte quella di coloro che viceversa hanno avuto l'intervento. Tale rapporto decresce nelle classi successive fino a non superare il valore di 1,5, ed è deducibile dalle tabelle presentate.

L'importanza espressa dall'intervento chirurgico durante il ricovero viene evidenziata dalle varie relazioni di condizionamento che questa variabile ha con i

motivi del ricovero, con i giorni intercorsi dalla richiesta del ricovero e con le modalità di prescrizione dello stesso (pronto soccorso, ospedale o altra modalità). Quest'ultima svolge un'azione filtro molto importante sulla scelta della sede di cura e sull'eventualità di un impegno di partecipazione alle spese di ospedalizzazione da parte del paziente e della sua famiglia. Nella rete risulta evidente infatti il legame diretto con le variabili che spiegano la sede e la natura dell'ospedale, oltre che con il contributo delle spese per il ricovero.

Il periodo di rilevazione e la ripartizione geografica risultano indipendenti nella rete: l'informazione che queste apportano alla comprensione del problema dell'assistenza informale non è catturata o è mediata da altre variabili.

La forza delle relazioni trovate fra le variabili che descrivono il contesto analizzato è specificata a livello quantitativo attraverso l'uso della probabilità. Attraverso le misure delle probabilità condizionate e del rischio relativo, è più facile capire come alcune dipendenze siano più forti rispetto ad altre. Per esempio, dalla Tab. 2, si può notare come la conoscenza dell'età del paziente, così come l'informazione sull'avvenuto intervento o meno, siano elementi abbastanza discriminatori per ottenere l'informazione sulla presenza di assistenza, variabile target dell'analisi, cosa che porta a supporre che la forza della relazioni tra queste variabili sia molto elevata. Ciò è confermato anche dal fatto che l'assenza di uno di questi due legami porta alla diminuzione della probabilità del modello (*score*) di un miliardo rispetto alla probabilità della rete in cui questi legami sono presenti.

5 Conclusioni

Nell'analisi delle relazioni tra componenti dell'assistenza informale intra- ed extra-ospedaliera espressa dai ricoverati emerge il contributo forte che alcune variabili – in modo particolare l'età e l'intervento chirurgico – svolgono nella costruzione delle relazioni.

I dati esaminati provengono da una serie di indagini campionarie condotte a livello nazionale presso le famiglie italiane sulle condizioni di salute e ricorso ai servizi socio-sanitari. Sfortunatamente, non sono disponibili dati su variabili oggettive quali la causa (clinica) del ricovero, delle dimissioni e il decorso post-ricovero in famiglia nonché informazioni sul personale e sulla struttura ospedaliera del ricovero. Queste informazioni avrebbero potuto integrare quanto dichiarato dagli intervistati e fornire elementi aggiuntivi alla conoscenza del fenomeno.

L'analisi svolta ha evidenziato la complessità dell'assistenza informale intra ed extra ospedaliera che rappresenta in Italia e in Europa un argomento emergente e destinato ad assumere sempre maggior interesse anche in considerazione della

crescente incidenza della popolazione anziana e del dibattito politico sulla riorganizzazione del *welfare*.

Riferimenti bibliografici

- ARMITAGE P., BERRY G. (1996) *Statistica medica: metodi statistici per la ricerca in medicina* (ed. italiana a cura di M. Bolzan), Mc Graw-Hill, Milano.
- BAYESWARE LIMITED (2000) On-line documentation: www.bayesware.com
- BOLZAN M. (1989) Condizione di salute e ricorso ad alcuni servizi sanitari della famiglia italiana. Analisi relativa alle coppie di coniugi, *II Congresso Nazionale Associazione Italiana di Statistica Medica e Programmazione Socio-Sanitaria*. Pavia
- BUNTINE W. (1994) Operations for Learning with Graphical Models, *Journal of Artificial Intelligence Research*, **2**: 159-225
- BUNTINE W. (1996a) Graphical Models for Discovering Knowledge. In: FAYYAD U.M. *et al.* (eds) *Advances in Knowledge Discovery and Data Mining*, AAAI Press: 59-82
- BUNTINE W. (1996b) A Guide on the Literature on Learning Graphical Models, *IEEE Transactions on Knowledge and Data Engineering*, **8**: 195-210
- COOPER G.F., HERSKOVITZ E. (1992) A Bayesian Method for the Induction of Probabilistic Networks from Data, *Machine Learning*, **9**: 309-347
- FAYYAD U.M., SHAPIRO G.P., SMYTH P., UTHURUSAMY R. (1996) *Advances in Knowledge Discovery and Data Mining*, AAAI Press
- HASTIE T., TIBSHIRANI R., FRIEDMAN J. (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer Verlag, New York
- HECKERMAN D. (1996) A Tutorial on Learning with Bayesian Networks. *Technical Report # MSR-TR-95-06*, Microsoft Research, Redmond, Washington
- HECKERMAN D., GEIGER D., CHICKERING D. (1995) Learning Bayesian Networks: the Combination of Knowledge and Statistical Data, *Machine Learning*, **20**: 197-243
- ISTAT (2001) *Rapporto Annuale. La situazione del Paese nel 2000*, ISTAT, Roma
- JENSEN F.V. (1996) *An Introduction to Bayesian Networks*, University College London Press, London
- LAURITZEN S.L. (1996) *Graphical Models*, Clarendon Press, Oxford
- LITMAN T.J. (1974) The Family as a Basic Unit in Health and Medical Care. A Socio-Behavioral Overview. *Social Science and Medicine*, **8**

- MITCHELL T. (1997) *Machine Learning*. Mc Graw-Hill, New York
- NEAPOLITAN R.E. (2004) *Learning Bayesian Networks*, Prentice Hall Series in Artificial Intelligence
- PEARL J. (1988) *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo, CA
- PLESS I.B.(1984) The Family as a Resource Unit In Health Care: Changing Patterns, *Social Science and Medicine*, **19(4)**: 385-389
- SINGH M., VALTORTA M. (1995) Construction of Bayesian Network Structures from Data: a Brief Survey and an Efficient Algorithm, *International Journal of Approximate Reasoning*, **12**
- SPIEGELHALTER D.J., LAURITZEN S.L. (1990) Sequential Updating of Conditional Probabilities on Directed Graphical Structures, *Networks*, **20(5)**: 579-605

The Exploratory Graphical Models for the Need of Hospital Informal Assistance Analysis

Summary. *The main aim in the researches in the social-health and economic fields is to search for determinants of the system and for relations within it. The Knowledge Discovery approach is an advanced contribution in this field of research. The Knowledge Discovery in Database (KDD) is the whole process of search new formal knowledge from a dataset containing information about the problem which we want to study. There are many way to represent the extract knowledge: in this paper we focus on the Bayesian Networks, which are graphical models of probabilistic dependencies. This explorative analysis want to highlights the complexity of informal assistance into and outside the hospital expressed by the hospitalised persons, and want to express the strong contribution of some variables in the relationships which are activated inside this systems.*

Keywords: *Knowledge Discovery in Database, Bayesian Networks, Learning, Informal Assistance.*

Stampato nel mese di Luglio 2005
presso la C.L.E.U.P. "Coop. Libreria Editrice Università di Padova"
Tipografia: Via Belzoni, 118/3 - Padova (Tel. 0498753496) tipografia@cleup.it
Redazione e Amm.ne: Via Belzoni, 118/3 - Padova (Tel. 049650261) redazione@cleup.it
Libreria: Via Gradenigo, 2 - Padova (Tel. 0498071998) libreria@cleup.it
www.cleup.it