Università degli Studi di Milano–Bicocca
Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea Magistrale in Informatica

Tesi di laurea magistrale

Continuos time
Bayesian Network Classifiers

Sottotitolo

Candidato:
Leonardo Di Donato
Matricola 744739

Relatore:
Prof. F. Antonio Stella

Correlatore:
Dott. Daniele Codecasa

Anno Accademico 2012–2013

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit.

— Oscar Wilde

Dedicato a tutti gli appassionati di LaTeX.

INDICE

ELENCO DELLE FIGURE

ELENCO DELLE TABELLE

SOMMARIO

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

ABSTRACT

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit.
Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis.
Curabitur dictum gravida mauris.

— Donald Ervin Knuth

RINGRAZIAMENTI

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Milano, luglio 2013 L.

INTRODUZIONE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit.

IL PRIMO CAPITOLO offre una visione d'insieme della storia di LaTeX e ne vengono presentate le idee di fondo.

IL SECONDO CAPITOLO offre una visione d'insieme della storia di LaTeX e ne vengono presentate le idee di fondo.

IL TERZO CAPITOLO spiega le operazioni, veramente semplici, per installare LaTeX sul proprio calcolatore.

IL QUARTO CAPITOLO descrive sinteticamente le principali norme tipografiche della lingua italiana, utili nella composizione di articoli, tesi o libri.

IL QUINTO CAPITOLO descrive sinteticamente le principali norme tipografiche della lingua italiana, utili nella composizione di articoli, tesi o libri.

IL SESTO CAPITOLO descrive sinteticamente le principali norme tipografiche della lingua italiana, utili nella composizione di articoli, tesi o libri.

1 INTRODUZIONE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2 CTBN

In questo capitolo si introducono i concetti fondamentali relativi alle Continuos time Bayesian Network (CTBN). Le CTBN sono un framework capace di modellare processi stocastici a tempo continuo e con spazio degli stati discreto.

Prima di affrontare tale argomento si presentano alcuni concetti propedeutici a questo lavoro di tesi: le Bayesian Network (BN) e i processi di Markov (sezione 2.1).

2.1 FONDAMENTI

Le Continuos time Bayesian Network utilizzano concetti e idee provenienti da teorie afferenti l'area statistica e del machine learning. Al fine di conferire alla discussione sulle CTBN un quadro iniziale completo ed esauriente, si presentano quindi gli aspetti di maggior rilievo di tali argomenti.

BAYESIAN NETWORK
Le Continuos time Bayesian Network utilizzano una rappresentazione strutturata dello spazio degli stati propria della teoria delle Bayesian Network. Ne ereditano perciò gli aspetti chiave (e. g. indipendenza condizionale) nonché l'insieme delle tecniche algoritmiche per l'apprendimento e l'inferenza.

PROCESSI DI MARKOV
Le Continuos time Bayesian Network descrivono la dinamica evolutiva di variabili casuali tramite un processo di Markov omogeneo costituito da un insieme di processi di Markov condizionali.

2.1.1 Bayesian Network

Una Bayesian Network è un modello grafico probabilistico costituito da un grafo aciclico orientato (DAG)[1]. I nodi di tale grafo rappresentano un insieme di variabili casuali mentre gli archi evidenziano le dipendenze (e le indipendenze) condizionali fra esse [Korb e Nicholson, 2011]. Una BN rappresenta la distribuzione di probabilità

1 Un grafo aciclico orientato (anche detto grafo aciclico diretto o digrafo aciclico) è un tipo di grafo che non presenta cicli diretti: comunque si scelga un vertice non è possibile tornare ad esso percorrendo gli archi del grafo.

congiunta del suo insieme di variabili casuali tramite la distribuzione di probabilità condizionale di ognuna di essa (si veda l'equazione 2). Le BN sono quindi modelli grafico probabilistici con cui è possibile modellare in modo probabilistico le relazioni causali tra variabili. Esse risultano molto utili nella rappresentazione e analisi di domini caratterizzati da incertezza. Sono infatti usate in svariate applicazioni di supporto alle decisioni, bioinformatica, biologia computazionale, data mining, information retrieval e classificazione.

Rappresentazione

Di seguito si fornisce la definizione formale delle Bayesian Network e si introducono i loro aspetti basilari.

Definizione 1 (Bayesian Network)**.** Una Bayesian Network $\mathcal{B}$ è una coppia $\mathcal{B} = (\mathcal{G}, \theta_{\mathcal{G}})$ costituita da:

- $\mathcal{G} = (\mathbf{V}(\mathcal{G}), \mathbf{A}(\mathcal{G}))$, un grafo aciclico orientato dove:
  - $\mathbf{V}(\mathcal{G}) = \{V_1, \ldots, V_n\}$ è l'insieme dei nodi, ognuno dei quali è associato ad una distribuzione di probabilità condizionale (CPD)[2]
  - $\mathbf{A}(\mathcal{G}) \subseteq \mathbf{V}(\mathcal{G}) \times \mathbf{V}(\mathcal{G})$ è l'insieme degli archi fra i nodi $\mathbf{V}(\mathcal{G})$
- $\theta_{\mathcal{G}}$, insieme delle CPD dei nodi che specifica $\mathbf{P}_{\mathcal{B}}$, la distribuzione di probabilità congiunta delle variabili casuali $\mathbf{X}_{\mathbf{V}(\mathcal{G})}$ a cui corrispondono i nodi $\mathbf{V}(\mathcal{G})$.

Osservazione 1.1. Ogni nodo di una BN è condizionalmente indipendente (si veda la definizione 2) dai suoi non–discendenti dati i suoi nodi genitori.

La CPD di ogni variabile casuale $X_i \in \mathbf{X}_{\mathbf{V}(\mathcal{G})}$ esprime i suoi valori di probabilità in funzione dei valori assunti da $Pa(X_i)$, notazione con cui denotiamo l'*insieme dei nodi genitori* per ogni nodo o variabile casuale.

Un arco da un nodo genitore verso un nodo figlio di $\mathcal{G}$ rappresenta una dipendenza diretta fra le corrispettive variabili casuali [si veda Russell e Norvig, 2003, sezione 14.1]. I nodi non direttamente connessi rappresentano variabili casuali condizionalmente indipendenti dagli altri nodi (per quanto riguarda il concetto di *indipendenza condizionale* si rimanda alla definizione 2).

Prima di procedere con la discussione si introduce la Chain Rule, proprietà fondamentale delle BN.

2 Nel caso di variabili causali discrete, le CPD sono rappresentabili come delle tabelle che contengono i valori di probabilità di un nodo in funzione di tutte le possibili configurazioni dei nodi genitori (cioè l'insieme dei nodi da cui parte un arco che punta al nodo di interesse). Tali tabelle sono spesso chiamate tabelle di probabilità condizionale (CPT).

Teorema 1 (Chain Rule)**.** *Dato un insieme di variabili casuali e una distribuzione di probabilità congiunta definita su di esse è possibile calcolare qualsiasi elemento di tale distribuzione tramite le distribuzioni di probabilità condizionale delle variabili casuali.*

Perciò, dato un insieme di variabili casuali $A_1, \ldots, A_n$ è possibile calcolare il valore di tale membro della distribuzione di probabilità congiunta applicando la definizione di probabilità condizionale:

$$\mathbf{P}(A_1, \ldots, A_n) = \mathbf{P}(A_n \,|\, A_{n-1}, \ldots, A_1) \cdot \mathbf{P}(A_{n-1}, \ldots, A_1)$$

Ripetendo tale processo per ogni termine finale si ottiene:

$$\mathbf{P}\Big(\bigcap_{k=1}^{n} A_k\Big) = \prod_{k=1}^{n} \mathbf{P}\Big(A_k \,|\, \bigcap_{j=1}^{k-1} A_j\Big) \tag{1}$$

Applicando l'equazione 1 alle Bayesian Network diciamo che la distribuzione di probabilità congiunta $\mathbf{P}_{\mathcal{B}}$ si *fattorizza* rispetto al grafo $\mathcal{G}$ se è possibile scrivere:

$$\mathbf{P}_{\mathcal{B}}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \,|\, Pa(X_i)). \tag{2}$$

L'equazione 2 esprime quindi la *proprietà di fattorizzazione* della distribuzione congiunta del modello grafico, detta *distribuzione di probabilità globale*, ed è ciò che permette di descriverla efficientemente in funzione delle distribuzioni condizionali dei nodi [Russell e Norvig, 2003, sezione 14.2], dette *distribuzioni di probabilità locali*. Questa proprietà contiene in sè il concetto di *proprietà di Markov* (si veda la definizione 3), il quale attesta che ogni nodo di una Bayesian Network dipende solo ed esclusivamente dai suoi nodi genitori [Korb e Nicholson, 2011, sezione 2.2.4]. Si noti inoltre, che le Bayesian Network richiedono (DAG, definizione 1) che la loro componente $\mathcal{G}$ non contenga cicli affinché possano rispettare tale proprietà [Russell e Norvig, 2003, sezione 14.1].

Poiché, come detto, una Bayesian Network stabilisce che ogni nodo, dati i suoi genitori, è *condizionalmente indipendente* da ogni altro nodo che non sia un suo discendente, di seguito introduciamo tale concetto formalmente.

Definizione 2 (Indipendenza condizionale)**.** Un evento A è *condizionalmente indipendente* da un evento B, data l'evidenza su un evento C, qualora la conoscenza di B non apporta alcuna variazione alla probabilità di A rispetto a quella conseguente alla conoscenza di C. Formalmente, ciò significa che:

$$\mathbf{P}(A, B \,|\, C) = \mathbf{P}(A \,|\, B, C) \cdot \mathbf{P}(B \,|\, C) = \mathbf{P}(A \,|\, C) \cdot \mathbf{P}(B \,|\, C).$$

Da cui segue che:

$$A \perp B \,|\, C \iff \mathbf{P}(A \,|\, B, C) = \mathbf{P}(A \,|\, C).$$

In termini non formali, supponendo di essere nel caso della definizione, cioè di avere una variabile casuale A *condizionalmente indipendente* da B dato C, ciò significa che è possibile ignorare B poiché essa non ha alcun riflesso sulla distribuzione condizionale di A quando sia noto l'evento C.

Si noti che il concetto appena espresso gioca un ruolo importante per i modelli probabilistici, quali sono le Bayesian Network, semplificando i calcoli richiesti per l'inferenza e l'apprendimento. Le Bayesian Network ereditano questi benefici dell'indipendenza condizionale come conseguenza della loro definizione (si veda l'osservazione 1.1). Infatti, la distribuzione condizionale di ogni variabile casuale X_i dipende solo ed esclusivamente dal valore dei suoi genitori, $Pa(X_i)$, mentre ignora completamente i valori dei nodi che non discendono da essa, $Nd(X_i)$.

Grazie alla definizione 2 è possibile esprimere in modo formale il concetto appena espresso per ogni nodo $X_i \in \mathbf{X}_{\mathbf{V}(\mathcal{G})}$:

$$\mathbf{P}(X_i \,|\, E, Pa(X_i)) = \mathbf{P}(X_i \,|\, Pa(X_i)) \; \forall\, E \in Nd(X_i),$$

dove $Nd(X_i)$ è l'insieme dei nodi non–discendenti (ed E è una variabile casuale o un insieme di variabili casuali ad essi associati). In base a ciò si dice quindi che le Bayesian Network rispettano l'*assunzione locale di Markov*.

Apprendimento e Inferenza

In questa sezione si descrivono brevemente e a scopo introduttivo i processi di apprendimento e inferenza sulle Bayesian Network.

Il problema dell'apprendimento per le Bayesian Network si divide principalmente in due casi:

- apprendere le CPD, nota la struttura
- apprendere sia le CPD, sia la struttura (incognita).

In entrambi i casi è di grande aiuto la rappresentazione efficiente delle Bayesian Network che, tramite la *fattorizzazione* della distribuzione di probabilità congiunta, permette di rappresentarla in modo compatto (tramite l'equazione 2) riducendo notevolmente il numero di parametri da calcolare.

Come detto, per specificare completamente una Bayesian Network è necessario rappresentare completamente la distribuzione di probabilità congiunta delle sue variabili tramite la distribuzione di probabilità condizionale di ognuna di esse. In generale, tali distribuzioni condizionali possono avere una qualsiasi forma anche se, al fine di semplificare i calcoli, è comune utilizzare distribuzioni discrete o Gaussiane per modellarle. Nel caso in cui i dati siano parzialmente osservabili solitamente si procede tramite l'algoritmo di Expectation

Maximization (EM), il quale alterna il calcolo dei valori attesi delle variabili casuali non osservate condizionalmente ai dati osservati con la massimizzazione della likelihood. Tale approccio generalmente converge ai valori di massima probabilità a posteriori per i parametri [si veda Dempster *et al.*, 1977].

Per l'*apprendimento dei parametri* esistono comunque una varietà di altri approcci possibili [si veda Heckerman, 1996] (e. g. trattare i parametri come variabili casuali sconosciute addizionali) che tuttavia non sono argomento di questo lavoro di tesi.

Si noti che le Bayesian Network non sono solamente un *modello discriminativo ma anche generativo* poiché possono essere utilizzate per soddisfare query arbitrarie, cioè per effettuare *inferenza probabilistica*: calcolare la distribuzione a posteriori di un insieme di variabili casuali data l'osservazione (evidenza) di altre (sfruttando il *teorema di Bayes*). In letteratura [si veda Heckerman, 1996] sono stati esplorati molti metodi di *inferenza esatta*, quali ad esempio l'eliminazione tramite integrazione o somma delle variabili non osservate che non fanno parte della query probabilistica o il metodo clique tree proprogation. Questi metodi, come gli altri presenti in letteratura, dati tutti i possibili alberi di decomposizione del grafo, sono esponenziali rispetto alla larghezza minore rilevata fra essi. Per quanto riguarda invece gli algoritmi di *inferenza approssimata* si citano due tra i più comuni: l'importance sampling [Shachter e Peot, 1990] e i metodi Markov Chain Monte Carlo (MCMC) (*Gibbs sampling*, *Metropolis sampling*, e *Hybrid Monte Carlo sampling*), basati sul campionamento stocastico [si veda Geman e Geman, 1984; Gilks *et al.*, 1996; MacKay, 1998].

Nel caso in cui non si disponga della struttura di una BN è richiesto l'*apprendimento strutturale*. Gli algoritmi per l'apprendimento strutturale delle Bayesian Network possono essere divisi in due famiglie.

ALGORITMI BASATI SU VINCOLI
: Algoritmi che apprendono la struttura del grafo analizzando le relazioni probabilistiche derivanti dalla proprietà di Markov tramite test di indipendenza condizionale e costruendo un grafo che soddisfi le proprietà di *d-separazione*[3] corrispondenti. I modelli risultanti sono spesso interpretati come *modelli causali* [Pearl, 1988].

ALGORITMI BASATI SU PUNTEGGIO
: Algoritmi che assegnano un punteggio (tramite una funzione di scoring) a tutte le strutture candidate e utilizzando tecniche di ottimizzazione cercano di raggiungere il punteggio massimo. Gli algoritmi di ricerca greedy (i. e. *golosi*, *aggressivi* o *avidi* a

3 Concetto di separazione direzionale tra insieme di nodi collegato al concetto di *indipendenza condizionale*. Ad esempio, quando un insieme di nodi $\mathbf{E}$ *d-separa* un insieme di nodi $\mathbf{X} = \{A, B\}$ allora A e B sono condizionalmente indipendenti dato $\mathbf{E}$.

seconda della traduzione preferita dall'inglese) sono la scelta più comune, tuttavia qualsiasi procedura di ricerca può essere usata.

Gli *algoritmi basati su vincoli* sono basati sull'algoritmo Inductive Causation (IC) di Verma e Pearl [1991] che fornisce un contesto teorico finalizzato all'apprendimento delle strutture dei modelli causali. L'algoritmo IC può essere riassunto nei tre passi successivi.

- Apprendimento dello scheletro (i. e. grafo non diretto) della rete. Poiché la ricerca esaustiva non è, nella maggior parte dei casi, computazionalmente realizzabile, tutti gli algoritmi di apprendimento restringono la ricerca al *Markov blanket*[4] di ogni nodo.
- Impostare la direzione degli archi che fanno parte di una *v-structure*[5].
- Impostare la direzione degli archi fra i nodi rimanenti affinché il vincolo di aciclicità sia rispettato.

Gli *algoritmi basati su punteggio* sono invece delle applicazioni dei vari algoritmi di ricerca euristica (e. g. *hill climbing*, *tabu search*, *best first search*, *simulated annealing*) che utilizzano una funzione di *scoring*. Solitamente la funzione di *scoring* utilizza la probabilità a posteriori della struttura in esame, dato l'insieme dei dati di apprendimento (i. e. training set), ed è *score-equivalent*, affinché reti che definiscono la stessa distribuzione di probabilità abbiano lo stesso score [Chickering, 2013]. Tuttavia, per quanto questi algoritmi siano utilizzati molto frequentemente, essi sono esponenziali rispetto al numero di nodi della struttura del grafo. Inoltre, qualora si utilizzi una strategia di ricerca locale, è possibile che l'algoritmo restituisca come risultato un minimo locale (per evitare questa situazione si ricorre spesso a metodi di ricerca globale quali il MCMC). Si fa notare che è possibile ridurre il tempo necessario richiesto per l'apprendimento strutturale fissando un numero massimo di genitori candidati e cercando esaustivamente in insiemi di tale cardinalità una struttura che massimizzi l'informazione mutua fra variabili [Heckerman *et al.*, 1995].

2.1.2 Processi di Markov

Sempre al fine di preparare la discussione delle Continuos time Bayesian Network si prosegue presentando alcuni concetti propedeutici relativi ai processi di Markov, una categoria di processi stocastici con assenza di memoria [Loève, 1978].

4 Il *Markov blanket* di un nodo A è un insieme composto dai nodi genitori di A, dai suoi nodi figli e da tutti i nodi che condividono un figlio con A.

5 Una *v-structure* è una tripla di nodi $X_j \rightarrow X_i \leftarrow X_k$ incidenti su una connessione convergente.

Definizione 3 (Proprietà di Markov)**.** Secondo la proprietà di Markov gli stati futuri di un processo stocastico sono indipendenti dagli stati passati, avendo evidenza sullo stato presente di tale processo.

Formalmente, un processo stocastico X gode di tale proprietà, se e solo se vale la seguente equazione [Loève, 1978]:

$$\mathbf{P}(X(t+\Delta t) \,|\, X(t), X(s)) = \mathbf{P}(X(t+\Delta t) \,|\, X(t)), \tag{3}$$

per ogni s, e t tali che $s < t < \infty$.

I modelli che rispettano tale proprietà sono detti modelli che rispettano l'*assunzione di Markov*.

Di conseguenza la distribuzione di probabilità condizionale degli stati futuri di un processo stocastico che gode di tale proprietà è indipendente dagli stati passati dato quello attuale.

In altri termini ciò indica che lo stato futuro di una variabile casuale è *condizionalmente indipendente* (si veda la definizione 2) dalla sequenza dei suoi stati passati, avendo evidenza sul suo stato presente.

Dalla proprietà di Markov deriva la definizione dei processi di Markov.

Definizione 4 (Processo di Markov)**.** Si definisce come processo di Markov un processo stocastico che gode della proprietà di Markov [Loève, 1978].

Definizione 5 (Catena di Markov)**.** Un processo di Markov che può assumere solo un numero finito di stati è solitamente definito come una *catena di Markov* [si veda Norris, 1998, p. 10].

Esistono due tipi di processi di Markov: omogenei e non. Si procede quindi fornendone le definizioni.

Definizione 6 (Processo di Markov omogeneo)**.** Un processo di Markov è detto *omogeneo* qualora $\mathbf{P}(X(t+\Delta t) \,|\, X(t))$ non dipenda dal tempo t. Affinché ciò sia vero, ponenedo $t = 0$, deve risultare che:

$$\mathbf{P}(X(t+\Delta t) \,|\, X(t)) = \mathbf{P}(X(\Delta t) \,|\, X(0)). \tag{4}$$

Data quindi una variabile casuale X e l'insieme delle sue istanziazioni $val(X) = \{x_1, \dots, x_J\}$, $X(t)$ è un processo di Markov *omogeneo, a tempo continuo e stati finiti* se e solo se la sua dinamica è definibile in termini di:

- una distribuzione di probabilità iniziale $\mathbf{P}^0_X$ su $val(X)$
- una matrice di intensità $\mathbf{Q}_X$.

Definizione 7 (Matrice di intensità)**.** Una matrice di intensità (IM), rappresenta un *modello di transizione Markoviano*:

$$\mathbf{Q}_X = \begin{bmatrix} -q_{x_1} & q_{x_1 x_2} & \cdots & q_{x_1 x_k} \\ q_{x_2 x_1} & -q_{x_2} & \cdots & q_{x_2 x_k} \\ \vdots & \vdots & \ddots & \vdots \\ q_{x_k x_1} & q_{x_k x_2} & \cdots & -q_{x_k} \end{bmatrix}$$

Lo scopo di un matrice di intensità è descrivere il comportamento transiente di X, un processo di Markov omogeneo.

Affinché $\mathbf{Q}_X$ sia una matrice di intensità valida, ogni sua riga deve sommare a 0:

$$q_{x_i} = \sum_{i \neq j} q_{x_i x_j} \quad \text{con} \quad q_{x_i}, q_{x_i x_j} > 0$$

Data quindi una matrice di intensità $\mathbf{Q}_X$ essa descrive il comportamento transiente di $X(t)$. Se $X(0) = x_i$ allora il processo di Markov omogeneo (e indicizzato dal tempo t) $X(t)$ rimarrà nello stato x_i una quantità di tempo *esponenzialmente distribuita* rispetto al parametro q_{x_i}. Di conseguenza la *funzione di densità* f e la corrispondente *funzione di ripartizione*[6] F sono:

$$\begin{aligned} f(t) &= q_{x_i} \exp(-q_{x_i} t), \quad t > 0 \\ F(t) &= 1 - \exp(-q_{x_i} t), \quad t \geqslant 0 \end{aligned} \tag{5}$$

Mentre gli elementi sulla diagonale, q_{x_i}, codificano una quantità che può essere interpretata come la «*probabilità istantanea*» che X abbandoni lo stato x_i, gli elementi non sulla diagonale, $q_{x_{ij}}$, esprimono l'*intensità di transizione* dallo stato x_i allo stato x_j.

Possiamo quindi calcolare:

- il *tempo atteso* di una transizione uscente dallo stato x_i

$$1/q_{x_i}$$

- la «*probabilità istantanea*» di transizione dallo stato x_i allo stato x_j

$$\theta_{x_i x_j} = q_{x_i x_j} / q_{x_i}.$$

Quindi una matrice di intensità induce una distribuzione di probabilità locale fattorizzata in due parti:

- q_{x_i}, che esprime quando avvengono le transizioni attraverso una *distribuzione di probabilità esponenziale*

6 Nel calcolo delle probabilità la funzione di ripartizione di una variabile casuale X a valori reali, anche nota come funzione di distribuzione cumulativa, è la funzione che associa a ciascun valore x la probabilità che X assuma valori minori o uguali ad x.

- $\theta_{x_{ij}}$, che esprime la *distribuzione di probabilità multinomiale* tra coppie di stati $i \neq j$.

Si osservi infine come la matrice $\mathbf{Q}_X$ fa in modo che X soddisfi la proprietà di Markov poiché il comportamento futuro di X è definito solamente in base al suo stato attuale (vale l'equazione 4).

Definizione 8 (Processo di Markov condizionale)**.** Un processo di Markov le cui intensità di transizione variano nel tempo non in funzione del tempo ma in funzione dei valori assunti ad ogni determinato istante t da un insieme di altre variabili, che evolvono anch'esse come dei processi di Markov, è detto essere un processo di Markov condizionale (o processo di Markov non omogeneo).

Assumendo quindi che una variabile casuale X evolva come un processo di Markov $X(t)$ e che la sua dinamica sia condizionata da un insieme di altre variabili casuali $Pa(X)$, anch'esse dei processi di Markov, possiamo definire per tale variabile casuale una matrice di intensità condizionale (CIM) $\mathbf{Q}_{X \mid Pa(X)}$.

Se specifichiamo una distribuzione di probabilità iniziale su X abbiamo così definito un processo di Markov il cui comportamento dipende dalle istanziazioni dei valori di $Pa(X)$.

Definizione 9 (Matrice di intensità condizionale)**.** Dato un insieme di processi di Markov $Pa(X)$, una matrice di intensità condizionale $\mathbf{Q}_{X \mid Pa(X)}$ è costituita da un insieme di matrici di intensità $\mathbf{Q}_{X \mid pa_i(x)}$, una per ogni diversa istanziazione $pa_i(x)$ di $Pa(X)$ [Stella e Amer, 2012]:

$$\mathbf{Q}_{X \mid Pa(X)} = \left\{ \mathbf{Q}_{X \mid pa_1(x)}, \mathbf{Q}_{X \mid pa_2(x)}, \dots, \mathbf{Q}_{X \mid pa_n(x)} \right\}.$$

Ogni matrice di intensità di $\mathbf{Q}_{X \mid Pa(X)}$ è del seguente tipo:

$$\mathbf{Q}_{X \mid pa_i(x)} = \begin{bmatrix} -q_{x_1}^{pa_i(x)} & q_{x_1 x_2}^{pa_i(x)} & \cdots & q_{x_1 x_k}^{pa_i(x)} \\ q_{x_2 x_1}^{pa_i(x)} & -q_{x_2}^{pa_i(x)} & \cdots & q_{x_2 x_k}^{pa_i(x)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{x_k x_1}^{pa_i(x)} & q_{x_k x_2}^{pa_i(x)} & \cdots & -q_{x_k}^{pa_i(x)} \end{bmatrix}.$$

Si noti infine come l'*ordine* di una matrice di intensità corrisponde a k, la cardinalità dell'insieme dei valori assunti da X.

2.2 DEFINIZIONI PRELIMINARI

Nelle precedenti sezioni sono stati illustrati i concetti che si pongono a fondamento delle Continuos time Bayesian Network:

- le Bayesian Network: utili a comprendere la rappresentazione strutturata dello spazio degli stati delle CTBN, l'utilizzo della nozione di indipendenza condizionale e le conseguenti tecniche di apprendimento e inferenza

- i processi di Markov, omogenei e non, al fine di introdurre le modalità di rappresentazione (qualitativa e quantitativa) delle CTBN.

Prima di presentare le Continuos time Bayesian Network come una collezione di processi di Markov a tempo continuo non omogenei e con spazio degli stati discreto [Nodelman, 2007], si forniscono alcune definizioni utili per il prosieguo della discussione.

Definizione 10 (Variabile di processo)**.** Una variabile di processo $\mathbf{X}$, anche detta Process Variable (PV) [Nodelman, 2007], è un insieme di processi di Markov a tempo continuo $X(t)$.

Definizione 11 (Traiettoria)**.** Istanziazione di un insieme di valori per $X(t)$ al variare di t.

Definizione 12 (J-time-segment)**.** Partizionamento di un intervallo temporale $[0, T)$ in J intervalli chiusi a sinistra:

$$[0, t_1)\,;\,[t_1, t_2)\,;\,\dots\,;\,[t_{J-1}, T)$$

Definizione 13 (J-evidence-stream)**.** Dato un J-time-segment composto da J intervalli temporali e una Variabile di processo $\mathbf{X}$ composta da N variabili casuali, un J-evidence-stream è l'insieme delle istanziazioni comuni $\mathbf{X} = \mathbf{x}$ associate ad ogni J-time-segment per ogni sottoinsieme delle variabili casuali. È denotato con $(\mathbf{X}^1 = \mathbf{x}^1, \mathbf{X}^2 = \mathbf{x}^2, \dots, \mathbf{X}^J = \mathbf{x}^J)$, o più concisamente $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)$.

Un J-evidence-stream $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)$ è detto essere *fully observed* (completamente osservato) se lo stato di tutte le variabili $X_n \in \mathbf{X}$ è conosciuto in tutto l'intervallo $[0, T)$. Viceversa, un J-evidence-stream è detto *partially observed* (parzialmente osservato) [Stella e Amer, 2012].

2.3 RAPPRESENTAZIONE

Una Continuos time Bayesian Network è un modello grafico in cui ogni nodo rappresenta una variabile casuale i cui stati evolvono in modo continuo bel tempo. Le dinamiche evolutive degli stati dei nodi sono governate e dipendono dal valore che gli stati dei nodi padre[7] assumono [Stella e Amer, 2012]. Quindi ogni nodo è un processo di Markov condizionale (si veda la definizione 8) a tempo continuo e spazio degli stati discreto.

Una CTBN è composta principalmente da due componenti:

7 Con il termine «nodo padre», o *parent node*, si intende un nodo il cui stato condiziona quello di un altro nodo del modello grafico.

- una distribuzione di probabilità iniziale
- le dinamiche che regolano l'evoluzione nel tempo continuo della distribuzione di probabilità

Più formalmente si definisce:

Definizione 14 (Continuos time Bayesian Network)**.** Data una variabile di processo $\mathbf{X}$, insieme di processi di Markov $X_1, X_2, \ldots, X_N$ a tempo continuo e con spazio degli stati finito $val(X_n) = \{x_1, \ldots, x_J\}$ (dove $n = 1, \ldots, N$), una CTBN $\mathcal{N}$ su $\mathbf{X}$ consiste di:

- una distribuzione di probabilità iniziale $\mathbf{P}^0_{\mathbf{X}}$ specificata come una Bayesian Network $\mathcal{B}$ su $\mathbf{X}$
- un modello di transizione a tempo continuo, specificato da:
  - un grafo $\mathcal{G}$, orientato e non necessariamente aciclico, composto dai nodi $X_1, X_2, \ldots, X_N$, ognuno dei quali possiede un insieme di genitori denotato da $Pa(X_n)$
  - una matrice di intensità condizionale $\mathbf{Q}_{X_n \mid Pa(X_n)}$ per ogni nodo $X_n \in \mathbf{X}$.

Per ogni variabile causale $X_n \in \mathbf{X}$ di $\mathcal{N}$ si ha quindi un insieme di modelli di probabilità locali: $\mathbf{Q}_{X_n \mid Pa(X_n)}$, la CIM di X_n, è infatti un insieme di modelli di transizione Markoviani la cui cardinalità è pari a quella dell'insieme delle diverse istanziazioni di $Pa(X_n)$.

Si riscontra, quindi, quanto già affermato in precedenza (si veda 2.1), cioè che una CTBN esprime la sua dinamica evolutiva globale tramite un unico processo di Markov omogeneo, costituito da un insieme di processi di Markov condizionali (un insieme di CIM e relative distribuzioni di probabilità iniziali).

Si noti che, diversamente dalle Bayesian Network, nelle Continuos time Bayesian Network gli archi fra i nodi rappresentano le dipendenze nel tempo. Per tale motivo è possibile che la componente $\mathcal{G}$ del modello di transizione continuo contenga dei cicli. Tra l'altro, come vedremo nel prosieguo, la mancanza di tale vincolo di aciclicità porta a notevoli vantaggi computazionali relativamente all'apprendimento della struttura di una CTBN dai dati.

2.4 APPRENDIMENTO

In questa sezione si argomenta sulla probabilità di un *insieme di dati completo* rispetto a una Continuos time Bayesian Network. A tal fine si mostra come una CTBN, essendo un modello esponenziale, possa essere decomposta in un aggregato di modelli di probabilità locali relativi alle singole variabili casuali e espressa in termini di *statistiche sufficienti* aggregate.

Si affronta infine il processo di apprendimento dei parametri delle Continuos time Bayesian Network da *dati completi*. I processi di apprendimento relativi a dati non completi sono tralasciati poiché non facenti parte degli argomenti di questo lavoro di tesi.

Definizione 15 (Insieme di dati completo)**.** Dato un insieme di variabili casuali, un insieme di dati $\mathcal{D} = \{\delta_1, \dots, \delta_h\}$ si dice *completo* se ogni δ_i (con $i = 1, \dots, h$) è un insieme di traiettorie completamente osservate delle variabili casuali (i. e. l'istanziazione di tutte le variabili casuali è osservabile per ogni istante temporale di ogni traiettoria).

2.4.1 Statistiche sufficienti

Le *statistiche sufficienti* per un singolo processo di Markov omogeneo $X(t)$ riassumono la sua dinamica evolutiva con:

- $T[x]$: la quantità di tempo trascorsa nello stato x
- $M[x, x']$: il numero di transizioni dallo stato x allo stato x'.

Il numero totale di transizioni uscenti da uno stato x è:

$$M[x] = \sum_{x'} M[x, x'].$$

Nel caso di un processo di Markov condizionale è invece necessario considerare anche l'istanziazione dell'insieme $Pa(X)$ dei nodi genitori:

- $T[x|pa_i(x)]$: la quantità di tempo trascorsa nello stato x quando $Pa(X) = pa_i(x)$
- $M[x, x'|pa_i(x)]$: il numero di transizioni dallo stato x allo stato x' quando $Pa(X) = pa_i(x)$.

Chiaramente, il numero totale di transizioni si calcola come sopra.

2.4.2 Likelihood

Al fine di presentare il calcolo della likelihood[8] di una CTBN rispetto a un dataset completo $\mathcal{D}$ è bene procedere per gradi e iniziare presentando dapprima la likelihood di una singola transizione di un singolo processo di Markov omogeneo $X(t)$.

Likelihood di una singola transizione

Data una tripla $d = \langle x_d, t_d, x_{d'} \rangle \in \mathcal{D}$, la quale esprime una transizione di $X(t)$ da x_d a $x_{d'}$ dopo che esso ha trascorso t_d tempo in

8 La likelihood di un insieme di valori per i parametri, dato un insieme di dati, è uguale alla probabilità dei dati, dati tali valori per i parametri.

x_d, è possibile scrivere la likelihood di questa singola transizione d in funzione dei parametri (2.4.3):

$$\begin{aligned} L_X(q\,,\theta : d) &= L_X(q : d)\, L_X(\theta : d) \\ &= q_{x_d}\, \exp(-q_{x_d}\, t_d)\, (\theta_{x_d x_{d'}}). \end{aligned} \tag{6}$$

Si noti che l'equazione 6 è ricavata moltiplicando la *funzione di distribuzione di probabilità* di $X(t)$ (equazione 5) per la *«probabilità istantanea»* di transizione (si veda la definizione 7).

Likelihood di un dataset completo

Poiché tutte le transizioni sono osservabili, la *likelihood* del dataset $\mathcal{D}$ può essere decomposta come un prodotto delle likelihood individuali di ogni singola transizione d [si veda Nodelman *et al.*, 2002, p. 3]. Per tale motivo $\mathcal{D}$ è sintentizzabile aggregando le *statistiche sufficienti* relative a ogni processo di Markov condizionale di una CTBN.

Quindi la likelihood di un dataset completo $\mathcal{D}$ rispetto a un singolo processo di Markov omogeneo $X(t)$ è:

$$\begin{aligned} L_X(q\,,\theta : \mathcal{D}) &= \Big(\prod_{d\in\mathcal{D}} L_X(q : d)\Big)\Big(\prod_{d\in\mathcal{D}} L_X(\theta : d)\Big) \\ &= \Big(\prod_{x} q_x^{M[x]} \exp(-q_x T[x])\Big)\Big(\prod_{x}\prod_{x\neq x'} \theta_{xx'}^{M[x,x']}\Big). \end{aligned} \tag{7}$$

Si supponga ora di traslare questo concetto a una Continuos time Bayesian Network $\mathcal{N}$ con N nodi: per ogni nodo X_i, con $i = 1,\dots N$ è necessario considerare tutte le transizioni contestualmente all'istanziazione dell'insieme $Pa(X_i)$ dei suoi nodi genitori. Poiché, nel caso di *dati completi*, si conosce sempre l'istanziazione di $Pa(X_i)$, allora, per ogni istante di tempo t, si conosce quale matrice di intensità $\mathbf{Q}_{X_i\,|\,pa_i(x)}$, con $pa_i(x) \in Pa(X_i)$, governi la dinamica di X_i.

Perciò la probabilità dei dati $\mathcal{D}$ rispetto a $\mathcal{N}$ è il prodotto delle likelihood di ogni variabile X_i:

$$\begin{aligned} L_{\mathcal{N}}(q\,,\theta\, : \mathcal{D}) &= \prod_{X_i\in\mathbf{X}} L_{X_i}(q_{X_i\,|\,Pa(X_i)}\,, \theta_{X_i\,|\,Pa(X_i)} : \mathcal{D}) \\ &= \prod_{X_i\in\mathbf{X}} L_{X_i}(q_{X_i\,|\,Pa(X_i)} : \mathcal{D})\, L_{X_i}(\theta_{X_i\,|\,Pa(X_i)} : \mathcal{D}). \end{aligned} \tag{8}$$

Il termine $L_X(\theta_{X\,|\,Pa(X)} : \mathcal{D})$ esprime la likelihood delle transizioni tra stati. Si osservi, inoltre, come il tempo che intercorre fra le transizioni sia trascurato poiché esse dipendono esclusivamente dal valore di nodi genitori [si veda Nodelman *et al.*, 2002, p. 3]. Quindi, usando le *statistiche sufficienti* si può scrivere:

$$L_X(\theta_{X\,|\,Pa(X)} : \mathcal{D}) = \prod_{pa_i(x)}\prod_{x}\prod_{x\neq x'} \theta_{xx'\,|\,pa_i(x)}^{M[x,x'\,|\,pa_i(x)]}.$$

Per quanto riguarda il calcolo di $L_X(q_{X\,|\,Pa(X)} : \mathcal{D})$ va considerato il caso in cui il tempo trascorso da X in uno determinato stato x termini non a causa di una sua transizione bensì a causa di una transizione di uno o più nodi appartenenti all'insieme dei suoi nodi genitori (i. e. una nuova istanziazione per l'insieme dei genitori $Pa(X)$). È quindi necessario considerare la probabilità che il nodo X rimanga in x una quantità di tempo *almeno pari* a t mentre i suoi nodi genitori $Pa(X)$ non effettuano alcuna transizione di stato [si veda Nodelman *et al.*, 2002, p. 3]. Tale quantità si ricava dalla funzione di distribuzione cumulativa di una distribuzione esponenziale (si veda l'equazione 5):

$$1 - F(t) = \exp(-q_{x\,|\,pa_i(x)} \cdot t).$$

Perciò la likelihood delle quantità di tempo trascorse in ogni stato è:

$$L_X(q_{X\,|\,Pa(X)} : \mathcal{D}) = \textstyle\prod_{pa_i(x)} \prod_x q_{x\,|\,pa_i(x)}^{M[x\,|\,pa_i(x)]} \exp(-q_{x\,|\,pa_i(x)}\, T[x\,|\,pa_i(x)]).$$

La likelihood di $\mathcal{N}$ è quindi:

$$\begin{aligned} L_{\mathcal{N}}(q\,,\,\theta : \mathcal{D}) = \prod_{pa_i(x)} \prod_{x} \Big(& q_{x\,|\,pa_i(x)}^{M[x\,|\,pa_i(x)]} \exp(-q_{x\,|\,pa_i(x)}\, T[x\,|\,pa_i(x)]) \cdot \\ & \cdot \prod_{x \neq x'} \theta_{xx'\,|\,pa_i(x)}^{M[x,x'\,|\,pa_i(x)]} \Big). \end{aligned} \tag{9}$$

Mentre, scrivendola come *log-likelihood*, si ottiene:

$$\begin{aligned} \ell_{\mathcal{N}}(q\,,\,\theta : \mathcal{D}) = \sum_{pa_i(x)} \sum_{x} \Big(& M[x\,|\,pa_i(x)] \ln(q_{x\,|\,pa_i(x)}) - q_{x\,|\,pa_i(x)} T[x\,|\,pa_i(x)] + \\ & + \sum_{x \neq x'} M[x\,,\,x'\,|\,pa_i(x)] \ln(\theta_{xx'\,|\,pa_i(x)}) \Big). \end{aligned} \tag{10}$$

In questa sezione si è presentato come computare la likelihood di un modello di una CTBN rispetto a un dataset completo.

Tuttavia, nel caso in cui non si conoscano i parametri di una CTBN è necessario stimarli. Nella prossima sezione viene affrontato esattamente questo argomento.

2.4.3 Stima dei parametri

Si affronta ora il problema dell'apprendimento dei parametri di una Continuos time Bayesian Network (con struttura nota $\mathcal{G}$) da un insieme di dati completi [si veda Nodelman, 2007, sezione 5.1].

In base a quanto attestato dalla definizione stessa delle CTBN (14), la dinamica evolutiva globale di una CTBN, cioè la dinamica di tutti i nodi di $\mathcal{G}$ (dei processi di Markov condizionali indicizzati dal tempo), è espressa tramite un processo di Markov omogeneo. Dalla definizione 7, inoltre, si deduce che tale processo di Markov induce un modello di probabilità composto da una *distribuzione esponenziale* con parametro $q_{x\,|\,pa_i(x)}$, che esprime il tempo trascorso in uno stato x da un nodo X data una istanziazione $pa_i(x)$ per i nodi genitori $Pa(X)$, e una *distribuzione multinomiale* con parametro $\theta_{xx'|pa_i(x)}$, che esprime il numero di transizioni uscenti da uno stato x verso x' (sempre fermo restando il condizionamento dato dall'istanziazione dei nodi genitori).

La media di tale distribuzione esponenziale è pari a $1/q_{x|pa_i(x)}$. Questa quantità esprime il tempo medio delle transizioni uscenti da uno stato x, fermo restando che il genitore del nodo in questione abbia istanziazione costante e uguale a $pa_i(x)$. Poiché il tempo medio si calcola rapportando il tempo totale trascorso in x, $T[x|pa_i(x)]$, rispetto al numero totale di transizioni uscenti da x, $M[x|pa_i(x)]$, si ottiene:

$$\frac{1}{q_{x|pa_i(x)}} = \frac{T[x|pa_i(x)]}{M[x|pa_i(x)]}.$$

Invece, la probabilità di transizione da uno stato x verso x' è data dal rapporto tra il numero totale di transizioni da x a x' diviso il numero totale di transizioni uscenti da x; cioè:

$$M[x, x'|pa_i(x)]/M[x|pa_i(x)].$$

Teorema 2. *Parametri maximum-likelihood (MLE). I parametri che massimizzano la likelihood dell'equazione 10 sono funzione delle statistiche sufficienti:*

$$\begin{aligned} q_{x|pa_i(x)} &= \frac{M[x|pa_i(x)]}{T[x|pa_i(x)]} \\ \theta_{xx'|pa_i(x)} &= \frac{M[x, x'|pa_i(x)]}{M[x|pa_i(x)]}. \end{aligned} \tag{11}$$

Si noti che, in questo caso (dataset completo), $q_{x|pa_i(x)}$ e $\theta_{xx'|pa_i(x)}$ sono delle stime esatte. Essi massimizzano la probabilità a posteriori di un dataset, dato un modello CTBN.

Da questi parametri è quindi possibile costruire le matrici di intensità condizionali di ogni nodo. Come si ricorderà, una CIM è un insieme di matrici di intensità, una per ogni istanziazione $pa_i(x)$ dei nodi genitori (si veda la definizione 9).

Perciò, fissato $pa_i(x)$, si può computare la rispettiva matrice di intensità per un nodo qualsiasi ponendo sulla diagonale il suo vettore dei parametri $q_{x|pa_i(x)}$ e ricavando i valori non sulla diagonale (i. e. la «*probabilità istantanea*» di transizione fra due stati; si veda a riguardo la definizione 7) dalla relazione fra i parametri q e θ:

$$q_{xx'|pa_i(x)} = \theta_{xx'|pa_i(x)} \cdot q_{x|pa_i(x)}. \tag{12}$$

3 CLASSIFICAZIONE

In questo capitolo viene introdotta una classe di modelli, che prende il nome di Continuos time Bayesian Network Classifier (CTBNC), il cui scopo è la classificazione supervisionata di traiettorie multivariate di variabili discrete a tempo continuo e spazio degli stati discreto. Si descrivono due istanze di tale classe: i classificatori Continuos time naïve Bayes (CTNB) nella sottosezione 3.1.1 e i classificatori Continuos time tree augumented naïve Bayes (CTTANB) nella sottosezione 3.1.2, per le quali si affronta il processo di apprendimento in caso di dati completi.

Infine, nella sezione 3.2, si presenta un algoritmo di inferenza esatta per la classe dei CTBNC.

3.1 APPRENDIMENTO

Ciao!

3.1.1 Naïve Bayes

...

3.1.2 Tree Augumented Naïve Bayes

...

3.2 INFERENZA

...

3.2.1 Naïve Bayes

...

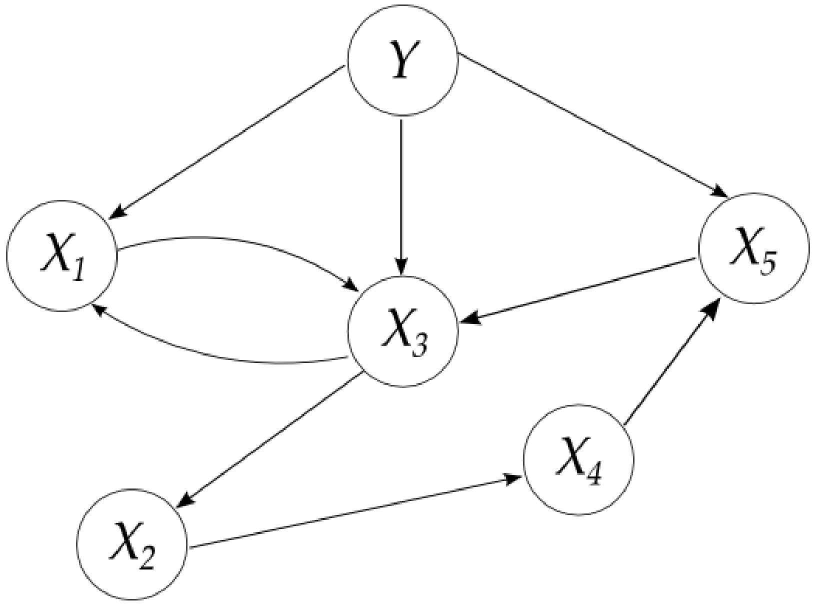

Figura 1: Un esempio di Continuos time Bayesian Network Classifier con cinque nodi attributo, X_1 , . . . , X_5, e un nodo classe, Y.

4 LEARNING STRUTTURALE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4.1 SCORE

...

4.2 RICERCA DELLA STRUTTURA

...

4.2.1 Hill Climbing

...

5 PACKAGE R

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5.1 ANALISI

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

5.2 PACKAGE CTBN

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus.

Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

6 CREAZIONE DI DATASET RELATIVI AL TRAFFICO

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

6.1 TSIS

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

6.1.1 Descrizione

...

6.1.2 API

...

6.2 estensione

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

6.2.1 Analisi

...

6.2.2 Sensors DLL

...

6.3 applicativi di supporto

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

7 ESPERIMENTI NUMERICI

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

7.1 DATASET 1

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

7.1.1 Modello TSIS

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

7.1.2 Risultati

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

7.2 DATASET 2

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

7.2.1 Modello TSIS

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

7.2.2 Risultati

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

8 CONCLUSIONI

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

A GUIDE ALL'USO

A.1 UTILIZZO DEL PACKAGE CTBN

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

A.1.1 Caricamento del dataset

A.1.2 Calcolo delle sufficient statistics

A.1.3 Calcolo dei parametri

A.1.4 Calcolo delle CIM

A.1.5 Apprendimento

A.1.6 Classificazione

A.1.7 Apprendimento strutturale

A.1.8 Cross-validation

A.2 CREAZIONE DI DATASET

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

A.2.1 Sensors DLL

...

Installazione

...

Guida all'uso

...

A.2.2 Applicativi di supporto

...

ACRONIMI

BIBLIOGRAFIA

Chickering, David Maxwell

2013 «A Transformational Characterization of Equivalent Bayesian Network Structures», *CoRR*, p. 87-98, `http://arxiv.org/abs/1302.4938`. (Citato a p. 7.)

Dempster, A P, N M Laird e D B Rubin

1977 «Maximum likelihood from incomplete data via the EM algorithm», *Journal of the Royal Statistical Society Series B Methodological*, Series B, 39, 1, p. 1-38, ISSN: 00359246, DOI: `10.2307/2984875`, `http://www.jstor.org/stable/2984875`. (Citato a p. 6.)

Geman, Stuart e Donald Geman

1984 «Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images», *IEEE Trans. Pattern Anal. Mach. Intell.*, 6, 6, p. 721-741, ISSN: 0162-8828, DOI: `10.1109/TPAMI.1984.4767596`, `http://dx.doi.org/10.1109/TPAMI.1984.4767596`. (Citato a p. 6.)

Gilks, WR, S Richardson e DJ Spiegelhalter

1996 «Markov chain Monte Carlo in practice». (Citato a p. 6.)

Heckerman, David

1996 «A Tutorial on Learning With Bayesian Networks», *Innovations in Bayesian Networks*, Studies in Computational Intelligence, 1995, November, a cura di Dawn E Holmes e Lakhmi C Jain, p. 33-82, ISSN: 1860949X, DOI: `10.1007/978-3-540-85066-3`, `http://www.springerlink.com/index/62mv333389016034.pdf`. (Citato a p. 6.)

Heckerman, David, Dan Geiger e David M. Chickering

1995 «Learning Bayesian networks: The combination of knowledge and statistical data», *Machine Learning*, 20, 3 [set. 1995], p. 197-243, ISSN: 0885-6125, DOI: `10.1007/BF00994016`, `http://link.springer.com/10.1007/BF00994016`. (Citato a p. 7.)

Korb, K.B. e A.E. Nicholson

2011 *Bayesian Artificial Intelligence*, Chapman & Hall / CRC Computer Science and Data Analysis, CRC PressINC, ISBN: 9781439815915. (Citato alle p. 2, 4.)

Loève, Michel

1978 *Probability theory. II Edition.* Fourth, Graduate Texts in Mathematics, Vol. 46, Springer-Verlag, New York, p. xvi+413, ISBN: 0-387-90262-7. (Citato alle p. 7, 8.)

MacKay, D. J. C.

1998 «Introduction to Monte Carlo methods», in *Proceedings of the NATO Advanced Study Institute on Learning in graphical models*, Kluwer Academic Publishers, Norwell, MA, USA, p. 175-204, http://dl.acm.org/citation.cfm?id=299068.299077. (Citato a p. 6.)

Nodelman, Uri, CR Shelton e Daphne Koller

2002 «Learning continuous time Bayesian networks», *Proceedings of the Nineteenth . . .*, X, arXiv:/arxiv.org/abs/1212.2498 [http:], http://dl.acm.org/citation.cfm?id=2100639. (Citato alle p. 14, 15.)

Nodelman, Uri D.

2007 *Continuos Time Bayesian Networks*, tesi di dott., Stanford University. (Citato alle p. 11, 15.)

Norris, James R.

1998 *Markov chains*, Cambridge series in statistical and probabilistic mathematics, Cambridge University Press, p. I-XVI, 1-237, ISBN: 978-0-521-48181-6. (Citato a p. 8.)

Pearl, Judea

1988 *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, ISBN: 0-934613-73-7. (Citato a p. 6.)

Russell, Stuart J. e Peter Norvig

2003 *Artificial Intelligence: A Modern Approach*, Pearson Education, ISBN: 0137903952, http://portal.acm.org/citation.cfm?id=773294. (Citato alle p. 3, 4.)

Shachter, Ross D. e Mark A. Peot

1990 «Simulation Approaches to General Probabilistic Inference on Belief Networks», in *Proceedings of the Fifth Annual Conference on Uncertainty in Artificial Intelligence*, UAI '89, North-Holland Publishing Co., Amsterdam, The Netherlands, p. 221-234, ISBN: 0-444-88738-5, http://dl.acm.org/citation.cfm?id=647232.719570. (Citato a p. 6.)

Stella, F e Y Amer

2012 «Continuous time Bayesian network classifiers.», *Journal of biomedical informatics*, 45, 6 [dic. 2012], p. 1108-19, ISSN: 1532-0480, DOI: 10.1016/j.jbi.2012.07.002, http://www.ncbi.nlm.nih.gov/pubmed/22846170. (Citato alle p. 10, 11.)

Verma, Thomas S. e Judea Pearl

1991 «Equivalence and synthesis of causal models», in *Uncertainty in Artificial Intelligence*, North Holland, p. 255-268. (Citato a p. 7.)

DICHIARAZIONE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices.

Milano, luglio 2013

Leonardo Di Donato