



Università degli Studi di Milano–Bicocca
Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea Magistrale in Informatica

Tesi di laurea magistrale

Continuous 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, consectetur adipiscing elit.

— Oscar Wilde

Dedicato a tutti gli appassionati di \LaTeX .

INDICE

1	INTRODUZIONE	1
2	CTBN	2
2.1	Fondamenti	2
2.1.1	Bayesian Network	2
2.1.2	Markov Process	6
2.2	Definizioni preliminari	9
2.3	Rappresentazione	10
2.4	Apprendimento	11
2.4.1	Sufficient Statistics	12
2.4.2	Likelihood	12
2.4.3	Stima dei parametri	14
2.4.4	Apprendimento	15
2.5	Inferenza	15
3	CLASSIFICAZIONE	16
3.1	Apprendimento	16
3.1.1	Naïve Bayes	16
3.1.2	Tree Augmented Naïve Bayes	16
3.2	Inferenza	16
3.2.1	Naïve Bayes	17
4	LEARNING STRUTTURALE	18
4.1	Score	18
4.2	Ricerca della struttura	18
4.2.1	Hill Climbing	18
5	PACKAGE R	19
5.1	Analisi	19
5.2	Package CTBN	19
6	CREAZIONE DI DATASET RELATIVI AL TRAFFICO	21
6.1	TSIS	21
6.1.1	Descrizione	21
6.1.2	API	21
6.2	Estensione	22
6.2.1	Analisi	22
6.2.2	Sensors DLL	22
6.3	Applicativi di supporto	22
7	ESPERIMENTI NUMERICI	23
7.1	Dataset 1	23

7.1.1	Modello TSIS	23
7.1.2	Risultati	24
7.2	Dataset 2	24
7.2.1	Modello TSIS	24
7.2.2	Risultati	24
8	CONCLUSIONI	26
A	GUIDE ALL'USO	27
A.1	Utilizzo del package CTBN	27
A.1.1	Caricamento del dataset	27
A.1.2	Calcolo delle sufficient statistics	27
A.1.3	Calcolo dei parametri	27
A.1.4	Calcolo delle CIM	27
A.1.5	Apprendimento	27
A.1.6	Classificazione	27
A.1.7	Apprendimento strutturale	27
A.1.8	Cross-validation	27
A.2	Creazione di dataset	27
A.2.1	Sensors DLL	28
A.2.2	Applicativi di supporto	28
	ACRONIMI	29
	BIBLIOGRAFIA	30

ELENCO DELLE FIGURE

ELENCO DELLE TABELLE

SOMMARIO

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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, consectetur 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, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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, consectetur 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.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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 delle Continuous time Bayesian Network (CTBN), un framework utile a modellare i processi stocastici relativi a uno spazio degli stati che evolve nel tempo. Tuttavia, prima di affrontare tale argomento si ritiene indispensabile presentare i suoi fondamenti matematici: le Bayesian Network (BN) e i Markov Process, trattati entrambi nella sezione 2.1.

2.1 FONDAMENTI

Come accennato, le Continuous time Bayesian Network sviluppano e integrano concetti e idee provenienti da altre 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 Continuous 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 efficienti per l'apprendimento e l'inferenza.

MARKOV PROCESS

Le Continuous time Bayesian Network descrivono la dinamica evolutiva delle variabili casuali da cui sono costituite tramite un Homogeneous Markov Process globale, costituito da un insieme di Conditional Markov Process.

2.1.1 Bayesian Network

Una Bayesian Network è un modello grafico probabilistico costituito da un grafo aciclico diretto (DAG).¹ I nodi di tale grafo rappresentano un insieme di variabili casuali mentre gli archi rappresentano le dipendenze (e le indipendenze) condizionali fra esse. Una BN rappresenta la distribuzione di probabilità congiunta del suo insieme di variabili casuali tramite la distribuzione di probabilità condizionale

¹ Un grafo aciclico diretto (anche detto grafo orientato aciclico 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.

di ognuna di essa 2. Le BN sono quindi modelli grafico probabilistici con cui è possibile modellare in modo probabilistico le relazioni causali dirette fra eventi. Perciò esse risultano molto utili nella rappresentazione e analisi di modelli del mondo reale che coinvolgono incertezza. Sono infatti usate in svariate applicazioni di supporto alle decisioni, bioinformatica, biologia computazionale, data mining, information retrieval e classificazione.

Rappresentazione

Di seguito si dà 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 diretto dove:
 - $\mathbf{V}(\mathcal{G}) = \{V_1, \dots, V_n\}$ è l'insieme dei nodi, ognuno dei quali è associato con una distribuzione di probabilità condizionale (CPD)²
 - ogni nodo è condizionalmente indipendente (2) dai suoi non-discendenti dati i suoi nodi genitori
 - $\mathbf{A}(\mathcal{G}) \subseteq \mathbf{V}(\mathcal{G}) \times \mathbf{V}(\mathcal{G})$ 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 $\mathbf{V}(\mathcal{G})$.

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 $\text{pa}(X_i)$, notazione con cui denotiamo l'insieme dei nodi genitori per ogni nodo o variabile casuale.

Mentre un arco da un nodo genitore verso un nodo figlio di \mathcal{G} rappresenta una dipendenza condizionale fra le corrispettive variabili casuali, i nodi non connessi, invece, rappresentano variabili casuali condizionalmente indipendenti dagli altri nodi (per quanto riguarda il concetto di *indipendenza condizionale* si rimanda invece a 2).

Prima di procedere con la discussione si introduce la Chain Rule, regola molto utile nel caso delle BN.

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 [3].

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).

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

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

Ripetendo tale processo per ogni termine finale si ottiene:

$$\mathbf{P}\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \mathbf{P}(A_k | \bigcap_{j=1}^{k-1} A_j) \quad (1)$$

Applicando la 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:

$$\mathbf{P}_{\mathcal{B}}(X_1, \dots, X_n) = \prod_{i=1}^n \mathbf{P}(X_i | \text{pa}(X_i)) \quad (2)$$

L'equazione 2 esprime quindi la *proprietà di fattorizzazione* della distribuzione congiunta del modello grafico, ed è ciò che permette di descriverla efficientemente in funzione delle distribuzioni condizionali dei nodi. Si noti inoltre, che le Bayesian Network richiedono che la loro componente \mathcal{G} non contenga cicli (si veda DAG in 1) affinché possano rispettare tale proprietà.

Poiché, come detto, una Bayesian Network stabilisce che ogni nodo, dati i suoi immediati ascendenti (i.e. nodi 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 .

Si noti che il concetto appena espresso gioca quindi un ruolo importante per i modelli probabilistici, quali sono le Bayesian Network: semplifica la struttura del modello e i calcoli richiesti per l'inferenza e l'apprendimento del modello. Le Bayesian Network ereditano questi

benefici dell'indipendenza condizionale come conseguenza della loro definizione (si veda 1): la distribuzione condizionale di ogni variabile casuale X_i dipende solo ed esclusivamente dal valore dei suoi genitori $pa(X_i)$ mentre ignora completamente le variabili casuali associate a nodi non-discendenti di X_i . In breve:

$$\forall X_i \in \mathbf{X}_{V(G)}:$$

$$P(X_i | E, pa(X_i)) = P(X_i | pa(X_i)) \quad \forall E \in nd(X_i)$$

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 solo 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 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. È frequente la situazione in cui queste distribuzioni condizionali includono parametri sconosciuti che è necessario stimare dai dati. In tali casi 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. Esistono comunque una varietà di altri approcci possibili (e.g. trattare i parametri come variabili casuali sconosciute addizionali) per l'*apprendimento dei parametri* 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 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 propagation. Questi metodi, come gli altri presenti in letterature, sono esponenziali rispetto al tree-width del grafo. Per quanto riguarda invece gli algoritmi di *inferenza approssimata* si citano due tra i più comuni: l'importance sampling e la simulazione Markov Chain Monte Carlo (MCMC).

Quando si dispone del grafo sottostante una BN, aldilà dell'osservabilità o meno delle sue variabili casuali, è possibile quindi usare tale BN per effettuare inferenza. Nel caso in cui invece non si disponga della struttura di una BN si presenta innanzitutto il problema dell'*apprendimento strutturale*. Per apprendere automaticamente il grafo di una Bayesian Network sono stati sviluppati principalmente due famiglie di algoritmi:

- algoritmi basati sulla ricerca dello scheletro del grafo dal quale, tramite le indipendenze condizionali osservate, si deducono successivamente le direzioni degli archi
- algoritmi che utilizzano tecniche di ricerca ottimizzata.

Gli algoritmi facenti parte della seconda famiglia utilizzando una strategia di ricerca (e.g. Hill Climbing, Best First Search, Simulated Annealing) e 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). Tuttavia, per quanto questi algoritmi siano utilizzati molto frequentemente, essi sono super 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.

2.1.2 Markov Process

Sempre al fine di preparare la discussione delle Continuous time Bayesian Network si prosegue presentando alcuni concetti propedeutici relativi ai Markov Process, una categoria di processi stocastici *memoryless* (con assenza di memoria).

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, se $X(t)$, $t > 0$ è un processo stocastico che gode di tale proprietà, allora $\forall h > 0$ vale la seguente equazione:

$$\mathbf{P}(X(t+h) | X(s) = x(s), s \leq t) = \mathbf{P}(X(t+h) | X(t) = x(t)) \quad (3)$$

Con la caratterizzazione *assunzione di Markov* si è soliti descrivere i modelli che rispettano tale proprietà.

Di conseguenza possiamo osservare che la distribuzione di probabilità condizionale degli stati futuri di un processo stocastico che gode di tale proprietà dipende solo ed esclusivamente dallo stato attuale del processo, data la conoscenza dei suoi stati passati.

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

Dalla proprietà di Markov deriva la definizione di Markov Process.

Definizione 4 (Markov Process). Si definisce come Markov process un processo stocastico che gode della proprietà di Markov.

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

Definizione 5 (Homogeneous Markov Process). Un processo di Markov è detto *omogeneo* qualora esso $\forall t, h > 0$ non dipenda dal tempo t , ovvero:

$$\mathbf{P}(X(t+h) | X(t) = x(t)) = \mathbf{P}(X(h) | X(0) = x(0)) \quad (4)$$

Data quindi una variabile casuale X e l'insieme delle sue istanziazioni $\text{val}(X) = \{x_1, \dots, x_I\}$, $X(t)$ 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}_X^0 su $\text{val}(X)$
- una matrice di intensità \mathbf{Q}_X .

Definizione 6 (Matrice di intensità). Una matrice di intensità, o Intensity Matrix (IM), rappresenta un *modello di transizione Markoviano*

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

il cui scopo è rappresentare e descrivere il comportamento transiente di X , un processo di Markov omogeneo.

Affinché \mathbf{Q}_X sia una Intensity Matrix valida, ogni sua riga deve sommare a 0. Ciò è possibile solo se:

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

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*³ F sono:

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

Mentre gli elementi sulla diagonale, q_{x_i} , codificano la *probabilità istantanea* che X abbandoni lo stato x_i , gli elementi non sulla diagonale⁴, 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

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

Osservazione 6.1. Si noti che poiché non è possibile che avvenga una transizione da x a x non esiste un parametro multinomiale $\theta_{x_i x_i}$.

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

- q_{x_i} , una *distribuzione esponenziale*, che esprime *quando* avvengono le transizioni
- $\theta_{x_{ij}}$, una *distribuzione multinomiale*, che esprime *verso quale stato* avvengono le transizioni.

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).

³ 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 .

⁴ Elementi della matrice per i cui indici i e j risulta vero che $i \neq j$.

Definizione 7 (Conditional Markov Process). 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 Conditional Markov Process (o Inhomogeneous Markov Process).

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 U , anch'esse dei processi di Markov, possiamo definire per tale variabile casuale una Conditional Intensity Matrix (CIM) $Q_{X|U}$.

Se specifichiamo una distribuzione di probabilità iniziale su X abbiamo così definito un Markov Process il cui comportamento dipende dalle istanziazioni dei valori di U .

Definizione 8 (Matrice di intensità condizionale). Dato un insieme di processi di Markov U , una matrice di intensità condizionale $Q_{X|U}$ è costituita da un insieme di matrici di intensità $Q_{X|u_i}$, una per ogni diversa istanziazione u_i di U .

$$Q_{X|U} = \{ Q_{X|u_1}, Q_{X|u_2}, \dots, Q_{X|u_n} \}$$

dove

$$Q_{X|u_i} = \begin{bmatrix} -q_{x_1}^{u_i} & q_{x_1 x_2}^{u_i} & \cdots & q_{x_1 x_k}^{u_i} \\ q_{x_2 x_1}^{u_i} & -q_{x_2}^{u_i} & \cdots & q_{x_2 x_k}^{u_i} \\ \vdots & \vdots & \ddots & \vdots \\ q_{x_k x_1}^{u_i} & q_{x_k x_2}^{u_i} & \cdots & -q_{x_k}^{u_i} \end{bmatrix}$$

2.2 DEFINIZIONI PRELIMINARI

Nelle precedenti sezioni sono stati illustrati i concetti che si pongono a fondamento delle Continuous 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 Markov Process, omogenei e non, al fine di introdurre le modalità di rappresentazione (qualitativa e quantitativa) esplicita delle dinamiche temporali.

È ora possibile quindi presentare le Continuous time Bayesian Network come una collezione di Continuous Time Markov Process (CTMP) non omogenei e con spazio degli stati discreto [1].

Definizione 9 (Continuous Time Markov Process). Variabile casuale $X(t)$, che gode della proprietà di Markov, indicizzata dal tempo $t \in [0, \infty)$.

Di seguito, invece, si riportano alcune definizioni che torneranno utili durante il prosieguo della discussione.

Definizione 10 (Process Variable). Una variabile di processo (PV) \mathbf{X} è un insieme di Continuous Time Markov Process $X(t)$.

Definizione 11 (Traiettoria). Istanziamento 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) [4].

2.3 RAPPRESENTAZIONE

Una Continuous time Bayesian Network è un modello grafico in cui ogni nodo è associato con una variabile casuale i cui stati evolvono nel tempo continuo. Le dinamiche evolutive degli stati dei nodi sono governate e dipendono dal valore che gli stati dei nodi padre⁵ assumono [4]. Quindi ogni nodo è un Conditional Markov Process (7) a tempo continuo e spazio degli stati discreto.

Una CTBN è composta principalmente da due componenti:

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

Più formalmente si definisce:

⁵ Con il termine «nodo padre», o *parent node*, si intende un nodo il cui stato condiziona quello di un altro nodo del modello grafico.

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

- una distribuzione di probabilità iniziale \mathbf{P}_X^0 specificata come una Bayesian Network \mathcal{B} su \mathbf{X}
- un modello di transizione continuo, specificato da:
 - un grafo \mathcal{G} , orientato e non necessariamente aciclico, composto dai nodi X_1, X_2, \dots, X_N , ognuno dei quali possiede un insieme di genitori denotato da $\text{pa}(X_n)$
 - una matrice di intensità condizionale $\mathbf{Q}_{X_n | \text{pa}(X_n)}$ 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: $\mathbf{Q}_{X_n | \text{pa}(X_n)}$, la CIM di X_n , è infatti un insieme di modelli di transizione Markoviani (definiti tramite delle matrici di intensità IM) la cui cardinalità è pari a quella dell'insieme delle diverse istanziazioni di $\text{pa}(X_n)$.

Si riscontra, quindi, quanto già affermato in precedenza (si veda 2.1), cioè che una CTBN, fissato un ordinamento delle variabili da cui è costituita, esprime la sua dinamica evolutiva globale tramite un unico Homogeneous Markov Process, costituito da un insieme di Conditional Markov Process (un insieme di CIM e relative distribuzioni di probabilità iniziali).

Si noti che, diversamente dalle Bayesian Network, nelle Continuous time Bayesian Network gli archi fra i nodi rappresentano le relazioni temporali fra essi. Tali relazioni codificano le dinamiche evolutive dei nodi, ognuna delle quali è espressa condizionatamente alla dinamica evolutiva degli stati dei suoi nodi genitori, tramite le matrici di intensità condizionale. 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 APPENDIMENTO

In questa sezione si argomenta circa la probabilità di un *insieme di dati completo*, data una Continuous time Bayesian Network. A tal fine si mostra quindi come essa possa essere decomposta in un aggregato di modelli di probabilità locali relativi alle singole variabili casuali e quindi espressa in termini di *sufficient statistics* aggregate, essendo le CTBN un modello esponenziale.

Si affronta infine il processo di apprendimento dei parametri delle Continuous 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). Un insieme $\mathcal{D} = \{\delta_1, \dots, \delta_n\}$ composto da una o più traiettorie di un insieme di variabili casuali. Ogni δ_i è un insieme completo (indicizzato dal tempo) di transizioni fra stati: l'istanziatura di tutte le variabili casuali è osservabile per ogni momento temporale di ogni traiettoria.

2.4.1 Sufficient Statistics

Le *sufficient statistics* per un singolo Homogeneous Markov Process $X(t)$ 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 Conditional Markov Process è invece necessario considerare anche l'istanziatura dell'insieme U dei nodi genitori:

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

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

2.4.2 Likelihood

Al fine di presentare il calcolo della likelihood⁶ 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 Homogeneous Markov Process $X(t)$.

⁶ Funzione di verosimiglianza di un evento, più debole della funzione di probabilità. Si applica solitamente ai parametri di un modello statistico. 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.

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

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

Si noti che l'equazione 6 è ricavata moltiplicando la *funzione di distribuzione di probabilità* di $X(t)$ (vedasi 5) per la *probabilità istantanea di transizione* (vedasi 6).

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 . Per tale motivo \mathcal{D} è sintetizzabile aggregando le *sufficient statistics* relative a ogni Conditional Markov Process di una CTBN.

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

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

Si supponga ora di traslare questo concetto a una Continuous 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'istanziamento dell'insieme \mathbf{U}_i dei suoi nodi genitori. Poiché, nel caso di *dati completi*, si conosce sempre l'istanziamento di \mathbf{U}_i , allora, per ogni preciso momento nel tempo t , si conosce quale matrice di intensità $\mathbf{Q}_{X_i | \mathbf{u}}$, con $\mathbf{u} \in \mathbf{U}_i$, 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 :

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

Il termine $L_X(\theta_{X | \mathbf{U}} : \mathcal{D})$ esprime la likelihood di tutte le sequenze di transizioni. Si osservi, inoltre, come il tempo che intercorre fra le transizioni sia trascurato poiché esse dipendono esclusivamente dal valore di nodi genitori. Quindi, usando le *sufficient statistics* si può scrivere:

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

Per quanto riguarda il calcolo di $L_X(q_{X|U} : \mathcal{D})$ va considerato il caso in cui il tempo trascorso in uno determinato stato x da un nodo qualsiasi X_i termini non a causa di un suo cambiamento di valore (transizione) ma 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 U_i). Ebbene, è quindi necessario includere la probabilità che X_i rimanga in x (mentre i nodi genitori non effettuano alcuna transizione di stato) una quantità di tempo almeno pari a t . Tale quantità si ricava dalla funzione di distribuzione cumulativa di una distribuzione esponenziale (si veda l'equazione 5):

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

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

$$L_X(q_{X|U} : \mathcal{D}) = \prod_u \prod_x q_{x|u}^{M[x|u]} \exp(-q_{x|u} T[x|u]).$$

La likelihood di \mathcal{N} è quindi:

$$L_N(q, \theta : \mathcal{D}) = \prod_u \prod_x \left(q_{x|u}^{M[x|u]} \exp(-q_{x|u} T[x|u]) \prod_{x' \neq x} \theta_{xx'|u}^{M[x, x'|u]} \right). \quad (9)$$

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

$$\ell_N(q, \theta : \mathcal{D}) = \sum_u \sum_x \left(M[x|u] \ln(q_{x|u}) - q_{x|u} T[x|u] + \sum_{x' \neq x} M[x, x'|u] \ln(\theta_{xx'|u}) \right). \quad (10)$$

In questa sezione si è presentato come computare la likelihood di un modello di una CTBN rispetto a un dataset completo; informazione che ha una chiara relazione, tramite il *teorema di Bayes*, con la probabilità di un dataset completo dato un modello di una CTBN.

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 approccia ora il problema dell'apprendimento dei parametri di una Continuous time Bayesian Network (con struttura nota \mathcal{G}) da un insieme di dati completi.

Sintetizzando le transizioni dei nodi di \mathcal{G} (che sono dei Conditional Markov Process indicizzati dal tempo) come un unico Homogeneous Markov Process, la rispettiva matrice di intensità (si veda 6) induce un modello di probabilità composto da una *distribuzione esponenziale* con parametro $q_{x|u}$, che esprime il tempo trascorso in uno stato x da un nodo data una istanziazione u per i nodi genitori, e una *distribuzione multinomiale* con parametro $\theta_{xx'|u}$, che esprime il numero di transizioni uscenti da uno stato x verso x' (sempre fermo restando il condizionamento dato dall'istanziazione dei nodi genitori).

Ebbene, la media della succitata distribuzione esponenziale è pari a $1/q_{x|u}$. Tale 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 u); cioè il tempo totale trascorso in x , $T[x|u]$, rapportato al numero totale di transizioni uscenti da x , $M[x|u]$. Perciò si ottiene:

$$\frac{1}{q_{x|u}} = \frac{T[x|u]}{M[x|u]}.$$

Invece, il numero medio di transizioni uscenti da uno stato x verso x' è dato dal rapporto tra il numero totale di transizioni da x a x' diviso il numero totale di transizioni uscenti da x ; cioè $M[x, x'|u]/M[x|u]$.

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

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

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

Da questi parametri (e indirettamente dalle sufficient statistics) è quindi possibile costruire le Conditional Intensity Matrix di ogni nodo. Come si ricorderà, una CIM è un insieme di matrici di intensità, una per ogni istanziazione u dei nodi genitori (si veda 8). Perciò, fissato u , si può computare la rispettiva matrice di intensità per un nodo qualsiasi ponendo sulla diagonale il suo vettore dei parametri $q_{x|u}$ e ricavando i valori non sulla diagonale dalla relazione fra i parametri q e θ (i.e. la probabilità istantanea di transizione fra due stati, 6):

$$q_{xx'|u} = \theta_{xx'|u} q_{x|u}.$$

2.4.4 Apprendimento

...

2.5 INFERENZA

...

3

CLASSIFICAZIONE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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.

3.1 APPENDIMENTO

...

3.1.1 Naïve Bayes

...

3.1.2 Tree Augmented Naïve Bayes

...

3.2 INFERENZA

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur 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.

3.2.1 Naïve Bayes

...

4

LEARNING STRUTTURALE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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, consectetur 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.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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, consectetur 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 consectetur.

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. Pellente-

sque 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, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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 UTILIZZO DEL PACKAGE CTBN

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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 ur-

na. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

A.2.1 Sensors DLL

...

Installazione

...

Guida all'uso

...

A.2.2 Applicativi di supporto

...

ACRONIMI

BN	Bayesian Network.....	2
CIM	Conditional Intensity Matrix.....	9
CPD	Conditional probability distribution.....	3
CTBN	Continuos time Bayesian Network	2
CTMP	Continuos Time Markov Process	9
DAG	Directed acyclic graph	2
EM	Expectation Maximization	5
IM	Intensity Matrix.....	7
MCMC	Markov Chain Monte Carlo.....	6
MLE	Maximum Likelihood Estimation	15
PV	Process Variable	10

BIBLIOGRAFIA

- [1] Uri Nodelman, CR Shelton e Daphne Koller. «Learning continuous time Bayesian networks». In: *Proceedings of the Nineteenth ... X* (2002). arXiv:/arxiv.org/abs/1212.2498 [http:]. URL: <http://dl.acm.org/citation.cfm?id=2100639> (cit. a p. 9).
- [2] Uri D. Nodelman. «Continuous Time Bayesian Networks». Tesi di dott. Stanford University, 2007.
- [3] Stuart J. Russell e Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2003. ISBN: 0137903952. URL: <http://portal.acm.org/citation.cfm?id=773294> (cit. a p. 3).
- [4] F Stella e Y Amer. «Continuous time Bayesian network classifiers.» In: *Journal of biomedical informatics* 45.6 (dic. 2012), pp. 1108–19. ISSN: 1532-0480. DOI: 10.1016/j.jbi.2012.07.002. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22846170> (cit. a p. 10).

DICHIARAZIONE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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