



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 Networks Classifiers

Sottotitolo

Candidato:
Leonardo Di Donato
Matricola 744739

Relatore:
Prof. F. Antonio Stella

Correlatore:
Dott. Daniele Codecasa

Anno Accademico 2012-2013

Questa tesi è stata scritta utilizzando L^AT_EX. Il modello tipografico che si è adottato è una personalizzazione dell'autore (reperibile all'indirizzo <https://github.com/leodido/arsclassica>) del modello offerto dal pacchetto ArsClassica. Si sono utilizzati i seguenti font:

- Palatino per il testo
- Euler per le equazioni, i listati e i numeri
- Iwona sia per le unità di sezionamento (capitoli, sezioni, sotto-sezioni) sia per le etichette degli elenchi di descrizioni e degli elementi fluttuanti.

Leonardo Di Donato: *Continuos Time Bayesian Networks Classifiers*.
Tesi di Laurea Magistrale © settembre 2013.

E-MAIL:

l.didonato@campus.unimib.it

leodidonato@gmail.com

Dedicato alla mia famiglia e a me stesso.

INDICE

1	CONTINUOUS TIME BAYESIAN NETWORK	1
1.1	Fondamenti	1
1.1.1	Bayesian Network	1
1.1.2	Processi di Markov	6
1.2	Definizioni preliminari	11
1.3	Rappresentazione	12
1.4	Apprendimento	13
1.4.1	Statistiche sufficienti	14
1.4.2	Likelihood	14
1.4.3	Stima dei parametri	16
2	CLASSIFICAZIONE	21
2.1	Modello	21
2.2	Apprendimento	24
2.3	Inferenza	28
3	APPRENDIMENTO STRUTTURALE	33
3.1	Funzione di scoring	34
3.2	Ricerca della struttura	36
3.2.1	Hill Climbing	37
4	PACKAGE R	40
4.1	R	40
4.2	Analisi	40
4.3	Package CTBN	40
4.3.1	Gestione dei dati	40
4.3.2	Apprendimento	40
4.3.3	Inferenza	40
4.3.4	Apprendimento strutturale	40
4.4	Cross-validazione	41
4.4.1	Metriche di valutazione	41
4.4.2	Package xvalidation	41
5	STRUMENTI PER LA CREAZIONE DI DATASET	42
5.1	TSIS	44
5.1.1	Componenti	45
5.1.2	Caratteristiche	47
5.2	Creazione di estensioni TSIS	50
5.2.1	Requisiti	51
5.2.2	Architettura di CORSIM	51
5.2.3	Ciclo di vita di CORSIM	53
5.2.4	Collegare una RTE a CORSIM	53

5.2.5	Utilizzo delle API	58
5.3	Estensione	60
5.3.1	Sensors DLL	60
5.3.2	Formato dell'output	63
5.4	Applicativi di supporto	65
6	ESPERIMENTI NUMERICI	66
6.1	Dataset #1	66
6.1.1	Modello TSIS	66
6.1.2	Risultati	71
6.2	Dataset #2	71
6.2.1	Modello TSIS	72
6.2.2	Risultati	80
7	CONCLUSIONI	82
A	GUIDE ALL'USO	83
A.1	Utilizzo del package CTBN	83
A.1.1	Caricamento del dataset	83
A.1.2	Calcolo delle sufficient statistics	83
A.1.3	Calcolo dei parametri	83
A.1.4	Calcolo delle CIM	83
A.1.5	Apprendimento	83
A.1.6	Classificazione	83
A.1.7	Apprendimento strutturale	83
A.1.8	Cross-validation	83
A.2	Creazione di dataset	83
A.2.1	Sensors DLL	83
A.2.2	Applicativi di supporto	83
	ACRONIMI	84
	INDICE ANALITICO	86
	BIBLIOGRAFIA	87

ELENCO DELLE FIGURE

Figura 2.1	Un esempio di CTBNC	23
Figura 2.2	Un CTNBC	23
Figura 2.3	Un CTTANBC	24
Figura 5.1	Onda quadra	43
Figura 5.2	Gestione del tempo in CORSIM	50
Figura 5.3	Diagramma dei componenti di CORSIM	52
Figura 5.4	Aggiunta di una RTE a TSIS	54
Figura 5.5	Barra degli strumenti di TShell	54
Figura 5.6	Collegamento delle funzioni della RTE a CORSIM	55
Figura 5.7	Configurazione delle proprietà di CORSIM	56
Figura 5.8	Diagramma delle classi di Sensors DLL	63
Figura 6.1	Rete stradale relativa al dataset #1	67
Figura 6.2	Piano semaforico relativo al dataset #1	68
Figura 6.4	Intersezioni della rete stradale del dataset #2	74
Figura 6.5	Rete stradale relativa al dataset #2	81

ELENCO DELLE TABELLE

Tabella 5.1	Durata dei passi temporali in FRESIM	49
Tabella 5.2	Ciclo di vita di CORSIM	57
Tabella 5.3	Semantica dell'output di Sensors DLL	64
Tabella 6.1	Sensori del dataset #1	69
Tabella 6.2	Periodi temporali relativi al dataset #1	69
Tabella 6.3	Flussi di ingresso relativi al dataset #1	70
Tabella 6.4	Percentuali di svolta relative al dataset #1	71
Tabella 6.5	Intersezioni relative al dataset #2	72
Tabella 6.6	Sensori relativi al dataset #2	76
Tabella 6.7	Periodi temporali del dataset #2	77
Tabella 6.8	Flussi di ingresso relativi al dataset #2	79

ELENCO DEGLI ALGORITMI

Algoritmo 2.1	Apprendimento di un classificatore CTNB	25
Algoritmo 2.2	Apprendimento di un classificatore CTBN	27
Algoritmo 2.3	Inferenza su un classificatore CTBN	31

Algoritmo 3.1	Algoritmo <i>hill climbing</i>	38
---------------	--	----

ELENCO DEI CODICI

Algoritmo 5.1	Costrutto per l'esportazione delle funzioni RTE	58
Algoritmo 5.2	Esempio di funzione RTE esportata	58
Algoritmo 5.3	Importazione delle CORWIN API	58
Algoritmo 5.4	Costrutto per l'importazione delle CORSIM API	59
Algoritmo 5.5	Importazione di oggetti delle CORSIM API . . .	59
Algoritmo 5.6	Rilevazione del passaggio dei veicoli sui sensori	62
Algoritmo 5.7	Formato di output di Sensors DLL	64
Algoritmo 6.1	Configurazione RNS del dataset #2	79

ABSTRACT

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.

*Le previsioni sono estremamente difficili.
Specialmente sul futuro.*

— Niels Bohr

RINGRAZIAMENTI

Desidero innanzitutto ringraziare il Prof. Stella per avermi concesso la possibilità di svolgere il mio periodo di tesi sotto la sua guida. Ciò mi ha permesso di accedere a un supporto, sia conoscitivo sia morale, che personalmente reputo esser stato di altissimo livello.

Porgo inoltre i miei ringraziamenti anche al correlatore, il Dott. Daniele Codecasa. I suoi suggerimenti e, in generale, gli scambi d'opinione intercorsi tra noi, hanno sicuramente aumentato il mio livello di preparazione e di conseguenza il livello di questo lavoro di tesi.

Infine intendo ringraziare vivamente la mia famiglia. Con il passare degli anni e l'aumentare della distanza mi rendo sempre più conto di quanto essi siano tutto ciò che ho e che voglio avere. Grazie.

Milano, settembre 2013

L.

INTRODUZIONE

Di seguito si passano in rassegna gli argomenti trattati in ogni capitolo.

NEL PRIMO CAPITOLO si introducono in primis i *fondamenti teorici* su cui il framework delle Continuous time Bayesian Network (CTBN) è basato. Successivamente si descrivono i *concetti* e gli *strumenti* da cui tale classe di modelli grafico probabilistici è costituita: *matrici di intensità condizionali* e *statistiche sufficienti*. Si affronta quindi il processo di *apprendimento dei parametri* delle CTBN da *dati completi* e, infine, il calcolo della *likelihood* di una CTBN rispetto a un insieme di *dati completi*.

NEL SECONDO CAPITOLO si affronta il problema della *classificazione supervisionata* di traiettorie multi-variate di variabili discrete a *tempo continuo*. A tal scopo si descrive una nuova classe di modelli, la classe dei classificatori Continuous time Bayesian Network (CTBNC), derivata dalle CTBN, e se ne presenta in particolare una sua specializzazione: i classificatori CTNB. Viene illustrato sia un algoritmo generale per l'*apprendimento dei parametri* dei classificatori CTBN, sia un algoritmo per l'*inferenza* della classe da associare a dei *dati completi* dato a un modello di classificatore CTNB (CTNBC) precedentemente appreso.

NEL TERZO CAPITOLO si affronta il problema dell'*apprendimento strutturale* di una CTBN da *dati completi*, per il quale si presenta un approccio risolutivo basato su punteggio. Viene presentata la definizione di una funzione che associa uno *score bayesiano* ad ogni struttura rispetto a dei dati di addestramento. Infine viene descritto il funzionamento di una *procedura di ottimizzazione* (i.e., nello specifico si descrive l'algoritmo *hill climbing*) finalizzata alla ricerca di una struttura che massimizzi la funzione di punteggio (i.e., *score bayesiano*).

NEL QUARTO CAPITOLO ...

NEL QUINTO CAPITOLO si introduce il problema di ottimizzazione del traffico urbano. Successivamente si presentano le caratteristiche e il funzionamento di Traffic Software Integrated System (TSIS), il sistema commerciale utilizzato per creare e simulare modelli di traffico. Infine si descrive Sensors DLL, estensione a tempo d'esecuzione di TSIS appositamente sviluppata al fine di monitorare e tracciare il passaggio dei veicoli sulle reti stradali tramite

sensori. Lo scopo di tale applicativo è la generazione di dataset sottoponibili agli algoritmi di classificazione e apprendimento strutturale delle CTBN. Tale passo è infatti propedeutico alla valutazione del processo di *classificazione dei profili di traffico* tramite classificatori CTBN.

NEL SESTO CAPITOLO si presentano le varie configurazioni dei 2 modelli di traffico, uno fittizio e uno che rispecchia una rete stradale reale (quella circostante *Viale C. Battisti, 20900 Monza, MB - Italia*), creati tramite TSIS e simulati tramite il relativo simulatore, CORSIM. Viene quindi fornita una descrizione dettagliata dei relativi dataset generati dai succitati modelli di traffico tramite Sensors DLL. Per tutti i dataset generati, vengono presentati e commentati i risultati ottenuti utilizzando i classificatori CTNB (CTNBC) per la *classificazione dei profili di traffico*. Si riportano, inoltre, i risultati dell'applicazione dell'algoritmo di *apprendimento strutturale* dei CTBN ai succitati dataset.

NEL SETTIMO CAPITOLO ...

L'APPENDICE A offre le *guide all'utilizzo* degli strumenti sviluppati a corredo di questo lavoro di tesi. In una prima parte viene presentato, tramite esempi, il funzionamento di RCTBN, il pacchetto R che implementa quanto trattato del framework CTBN. Nella seconda ed ultima parte viene invece presentata la generazione di dataset tramite l'esecuzione di Sensors DLL su modelli di traffico TSIS.

In questo capitolo si introducono i concetti fondamentali relativi alle Continuous 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 1.1).

1.1 FONDAMENTI

Le Continuous 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 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 per l'apprendimento e l'inferenza.

PROCESSI DI MARKOV

Le Continuous time Bayesian Network descrivono la dinamica evolutiva di variabili casuali tramite un costituito da un insieme di processi di Markov condizionali.

1.1.1 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 congiunta del suo insieme di variabili casuali tramite la distribuzione di probabilità condizionale di ognuna di essa (si veda l'equazione 1.2). Le

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

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 1.1 (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:
 - $\mathbf{V}(\mathcal{G}) = \{V_1, \dots, V_n\}$ 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 1.1.1. Ogni nodo di una BN è condizionalmente indipendente (si veda definizione 1.2) 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 $\text{Pa}(X_i)$, notazione con cui si denota 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 1.2).

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

Teorema 1.1 (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.*

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).

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:

$$\mathbf{P}\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \mathbf{P}\left(A_k \mid \bigcap_{j=1}^{k-1} A_j\right). \quad (1.1)$$

Applicando l'equazione 1.1 alle Bayesian Network si dice che la distribuzione di probabilità congiunta $\mathbf{P}_{\mathcal{B}}$ si *fattorizza* rispetto al grafo \mathcal{G} se è possibile scrivere:

$$\mathbf{P}_{\mathcal{B}}(X_1, \dots, X_n) = \prod_{i=1}^n \mathbf{P}(X_i | \text{Pa}(X_i)). \quad (1.2)$$

L'equazione 1.2 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 sé il concetto di *proprietà di Markov* (si veda la definizione 1.3), il quale attesta che ogni nodo di una Bayesian Network dipende solo ed esclusivamente dai suoi nodi genitori (Korb e Nicholson, 2011, sottosezione 2.2.4). Si noti inoltre, che le Bayesian Network richiedono (DAG, definizione 1.1) 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 si introduce tale concetto formalmente.

Definizione 1.2 (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 1.1.1). 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 1.2 è possibile esprimere in modo formale il concetto appena espresso per ogni nodo $X_i \in \mathbf{X}_{V(G)}$:

$$P(X_i | E, Pa(X_i)) = P(X_i | Pa(X_i)) \quad \forall E \in Nd(X_i),$$

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 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 1.2) 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 propagation*. Questi metodi, come gli altri presenti in letteratura, sono sempre esponenziali rispetto al *tree-width*³ del grafo. 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 S. Geman e D. 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*⁴ 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* 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 teo-

³ In teoria dei grafi, il *tree-width* è un numero associato ad un grafo. Esso corrisponde alla lunghezza minima di tutti i possibili alberi di decomposizione del grafo in esame. La lunghezza di un albero di decomposizione corrisponde alla dimensione massima dei suoi nodi, cioè sottoinsiemi dell'insieme dei vertici del grafo, sottratto 1.

⁴ Concetto di separazione direzionale tra insiemi di nodi collegato al concetto di *indipendenza condizionale*. Ad esempio, quando un insieme di nodi **E** *d-separa* un insieme di nodi **X** = {A, B} allora A e B sono condizionalmente indipendenti dato E.

rico 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*⁵ di ogni nodo.
- Impostare la direzione degli archi che fanno parte di una *v-structure*⁶.
- 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* è basata sulla *likelihood*, ovvero sulla la probabilità a posteriori dell'insieme dei dati di apprendimento (i. e., *training set*), data la struttura in esame e i parametri del modello. Tale funzione è spesso *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 probabile che l'algoritmo restituisca come risultato un minimo locale. 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).

1.1.2 Processi di Markov

Sempre al fine di preparare la discussione delle Continuous 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).

Definizione 1.3 (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.

⁵ 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 .

⁶ Una *v-structure* è una tripla di nodi $X_j \rightarrow X_i \leftarrow X_k$ incidenti su una connessione convergente.

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

$$\mathbf{P}(X(t + \Delta t) | X(t), X(s)) = \mathbf{P}(X(t + \Delta t) | X(t)), \quad (1.3)$$

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 1.2) dalla sequenza dei suoi stati passati, avendo evidenza sul suo stato presente.

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

Definizione 1.4 (Processo di Markov). Si definisce (Loève, 1978) come processo di Markov un processo stocastico che gode della proprietà di Markov.

Definizione 1.5 (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 1.6 (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 deve risultare che:

$$\mathbf{P}(X(t + \Delta t) | X(t)) = \mathbf{P}(X(\Delta t) | X(0)). \quad (1.4)$$

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

Definizione 1.7 (Matrice di intensità). Una matrice di intensità (IM), rappresenta un *modello di transizione Markoviano*:

$$\mathbf{Q}_X = \begin{bmatrix} -\mathbf{q}_{x_1} & \mathbf{q}_{x_1 x_2} & \cdots & \mathbf{q}_{x_1 x_K} \\ \mathbf{q}_{x_2 x_1} & -\mathbf{q}_{x_2} & \cdots & \mathbf{q}_{x_2 x_K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_{x_K x_1} & \mathbf{q}_{x_K x_2} & \cdots & -\mathbf{q}_{x_K} \end{bmatrix}.$$

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

Osservazione 1.7.1. L'ordine di una matrice di intensità corrisponde a $K = |\text{val}(X)|$, la cardinalità dell'insieme dei valori assunti da X .

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

$$\mathbf{q}_{x_i} = \sum_{i \neq j} \mathbf{q}_{x_i x_j} \quad \text{con} \quad \mathbf{q}_{x_i}, \mathbf{q}_{x_i x_j} > 0.$$

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 \mathbf{q}_{x_i} . Di conseguenza la *funzione di densità* f e la corrispondente *funzione di ripartizione*⁷ F sono quelle della distribuzione esponenziale:

$$\begin{aligned} f(t) &= \mathbf{q}_{x_i} e^{-\mathbf{q}_{x_i} t}, \quad t > 0 \\ F(t) &= 1 - e^{-\mathbf{q}_{x_i} t}, \quad t \geq 0. \end{aligned} \tag{1.5}$$

Quando un modello di transizione è definito esclusivamente tramite una matrice di intensità \mathbf{Q}_X si dice che esso usa una *parametrizzazione pura* delle intensità. In tal caso i parametri per un processo di Markov omogeneo con K stati sono $\{\mathbf{q}_{x_i}, \mathbf{q}_{x_i x_j} : 1 \leq i, j \leq K, i \neq j\}$.

Mentre gli elementi sulla diagonale di una matrice di intensità, \mathbf{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, $\mathbf{q}_{x_i x_j}$, esprimono l'*intensità di transizione* dallo stato x_i allo stato x_j .

Tuttavia, questa non è l'unica parametrizzazione possibile per un processo di Markov omogeneo. Si noti infatti che la distribuzione di probabilità locale sulle transizioni di X è fattorizzata in due parti.

Definizione 1.8 (Parametrizzazione mista delle intensità). La *parametrizzazione mista* delle intensità per un processo di Markov omogeneo X con K stati è composta da due insiemi di parametri:

$$\begin{aligned} \mathbf{q}_X &= \{\mathbf{q}_{x_i} : 1 \leq i \leq K\}, \\ \boldsymbol{\theta}_X &= \{\boldsymbol{\theta}_{x_i x_j} : 1 \leq i, j \leq K, i \neq j\}. \end{aligned}$$

La semantica di tali insiemi di parametri è la seguente:

- \mathbf{q}_X è un insieme di intensità \mathbf{q}_{x_i} che parametrizzano una *distribuzione di probabilità esponenziale* ed esprimono quando avvengono le transizioni

⁷ 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 (i.e., $\mathbf{P}(X \leq x)$).

- θ_X è un insieme di probabilità $\theta_{x_i x_j}$ che rappresentano la *probabilità di transitare* dallo stato x_i allo stato x_j , con $i \neq j$, sapendo che avverrà un salto ad un determinato istante di tempo.

Osservazione 1.8.1. Si osservi che, al di là del tipo di parametrizzazione con cui si sceglie di definire un modello di transizione, il numero di parametri necessari è pari a K^2 sebbene il numero di parametri liberi sia solo $K^2 - K$ (Nodelman, 2007).

Osservazione 1.8.2. Si noti, inoltre, che una parametrizzazione può essere più chiara dell'altra a seconda del processo in cui si è coinvolti, di conseguenza nel prosieguo le si utilizzerà entrambe in modo intercambiabile.

Al fine di correlare questi due tipi di parametrizzazione dei modelli di transizione si riporta il seguente teorema (Nodelman, 2007).

Teorema 1.2. *Dati X e Y , due processi di Markov omogenei con lo stesso spazio degli stati e la stessa distribuzione di probabilità iniziale, se il modello di transizione di X è definito tramite la matrice di intensità Q_X e quello di Y è definito tramite la parametrizzazione mista q_Y, θ_Y , allora X e Y sono stocasticamente equivalenti⁸ solo se:*

$$q_{y_i} = q_{x_i}$$

e

$$\theta_{y_i y_j} = \frac{q_{x_i x_j}}{q_{x_i}}.$$

Osservazione 1.8.3. Si osservi che il teorema 1.2 formalizza la relazione che sussiste fra i parametri q e θ .

Quindi, qualsiasi sia la parametrizzazione utilizzata per rappresentare il modello di transizione di un processo di Markov omogeneo X , è possibile 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 sapendo che avverrà un salto ad un determinato istante di tempo

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

Infine, si noti che la matrice 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 1.4).

⁸ Due processi di Markov sono detti *stocasticamente equivalenti* se posseggono lo stesso spazio degli stati e le stesse probabilità di transizione (Gihman e Skorohod, 1973).

Definizione 1.9 (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 $P_a(X)$, anch'esse dei processi di Markov, è possibile definire per tale variabile casuale una matrice di intensità condizionale (CIM) $\mathbf{Q}_{X|P_a(X)}$.

Specificando una distribuzione di probabilità iniziale su X si definisce quindi un processo di Markov il cui comportamento dipende dalle istanziazioni dei valori di $P_a(X)$.

Definizione 1.10 (Matrice di intensità condizionale). Dato un insieme di processi di Markov $P_a(X)$, una matrice di intensità condizionale $\mathbf{Q}_{X|P_a(X)}$ è costituita da un insieme di matrici di intensità $\mathbf{Q}_{X|p_{a_i}(x)}$, una per ogni diversa istanziazione $p_{a_i}(x)$ di $P_a(X)$ (Stella e Amer, 2012):

$$\mathbf{Q}_{X|P_a(X)} = \{ \mathbf{Q}_{X|p_{a_1}(x)}, \mathbf{Q}_{X|p_{a_2}(x)}, \dots, \mathbf{Q}_{X|p_{a_n}(x)} \}.$$

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

$$\mathbf{Q}_{X|p_{a_i}(x)} = \begin{bmatrix} -\mathbf{q}_{x_1}^{p_{a_i}(x)} & \mathbf{q}_{x_1 x_2}^{p_{a_i}(x)} & \dots & \mathbf{q}_{x_1 x_K}^{p_{a_i}(x)} \\ \mathbf{q}_{x_2 x_1}^{p_{a_i}(x)} & -\mathbf{q}_{x_2}^{p_{a_i}(x)} & \dots & \mathbf{q}_{x_2 x_K}^{p_{a_i}(x)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_{x_K x_1}^{p_{a_i}(x)} & \mathbf{q}_{x_K x_2}^{p_{a_i}(x)} & \dots & -\mathbf{q}_{x_K}^{p_{a_i}(x)} \end{bmatrix}.$$

Di seguito si presenta un breve esempio finalizzato alla comprensione pratica delle matrici di intensità condizionali (CIM) e del loro scopo.

Esempio 1.10.1.

Date due variabili causali, $E(t)$ e $H(t)$, delle quali la prima modella l'eventualità che un individuo stia mangiando o meno (se $e = 2$ allora l'individuo sta mangiando, viceversa se $e = 1$) mentre la seconda modella l'eventualità che lo stesso individuo abbia fame o meno (se $h = 2$ allora l'individuo è affamato, viceversa se $h = 1$) e la matrice di intensità condizionale $\mathbf{Q}_{E|H}$, che è un insieme composto dalle matrici di intensità $\mathbf{Q}_{E|h=1}$ e $\mathbf{Q}_{E|h=2}$, è possibile calcolare la probabilità degli eventi della variabile casuale E condizionatamente all'evidenza che si possiede sulla variabile casuale H .

$$\mathbf{Q}_{E|h=1} = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{bmatrix} -.01 & .01 \\ 10 & -10 \end{bmatrix} \end{matrix} \quad \mathbf{Q}_{E|h=2} = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{bmatrix} -2 & 2 \\ .01 & -.01 \end{bmatrix} \end{matrix}.$$

Ipotizzando che l'unità temporale corrisponda a un'ora:

- un individuo affamato ($h = 2$) che non sta mangiando ($e = 1$) inizierà a mangiare in 30 minuti poiché

$$\frac{1}{q_{e=1|h=2}} = \frac{1}{2},$$

- un individuo non affamato ($h = 1$) che sta mangiando ($e = 2$) smetterà di mangiare ($e = 1$) entro 6 minuti poiché

$$\frac{1}{q_{e=2|h=1}} = \frac{1}{10};$$

si osservi che la “*probabilità istantanea*” di transizione da $e = 2$ a $e = 1$ è

$$\theta_{e=2,e=1|h=1} = \frac{q_{e=2,e=1|h=1}}{q_{e=2|h=1}} = \frac{10}{10} = 1,$$

ciò poiché le matrici di intensità hanno dimensione 2×2 e, dovendo ogni loro riga sommare a 0, gli elementi sulla diagonale sono uguali al rispettivo (stessa riga) e unico elemento non sulla diagonale.

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

Prima di presentare le Continuous 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 1.11 (Variabile di processo). Una variabile di processo X , anche detta Process Variable (PV) (Nodelman, 2007), è un insieme di processi di Markov a tempo continuo $X(t)$.

Definizione 1.12 (Traiettoria). Istanziamento di un insieme di valori per $X(t)$ al variare di t .

Definizione 1.13 (J-time-segment). Partizionamento di un intervallo temporale $[0, T)$ in J intervalli chiusi a sinistra:

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

Nota 1.13.1. È possibile riferirsi a tale concetto anche tramite l'espressione "insieme dei segmenti temporali".

Definizione 1.14 (J-evidence-stream). Data una variabile di processo \mathbf{X} composta da N variabili casuali e un insieme di segmenti temporali composto da J intervalli, un *J-evidence-stream* è l'insieme delle istanziazioni comuni $\mathbf{X} = \mathbf{x}$ associate ad ogni intervallo temporale per ogni sottoinsieme delle variabili casuali (Stella e Amer, 2012). 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 con $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)$.

Nota 1.14.1. È possibile riferirsi a tale concetto anche tramite l'espressione "flusso di evidenze".

Nota 1.14.2. Un flusso di evidenze $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)$ è detto essere *completamente osservato* se lo stato di tutte le variabili $X_n \in \mathbf{X}$ è conosciuto in tutto l'intervallo $[0, T)$. Viceversa, un flusso di evidenze è detto *parzialmente osservato*.

1.3 RAPPRESENTAZIONE

Una Continuous time Bayesian Network è un modello grafico in cui ogni nodo rappresenta una variabile casuale i cui stati evolvono in modo continuo nel tempo. Le dinamiche evolutive degli stati dei nodi sono governate e dipendono dal valore che gli stati dei nodi padre⁹ assumono (Nodelman, 2007). Quindi ogni nodo è un processo di Markov condizionale (si veda la definizione 1.9) a tempo continuo e spazio degli stati discreto.

Una CTBN è composta principalmente da due componenti:

- una distribuzione di probabilità iniziale
- le componenti che regolano l'evoluzione nel tempo del sistema

Più formalmente si definisce:

Definizione 1.15 (Continuous time Bayesian Network). Data una variabile di processo \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 $\mathbf{P}_{\mathbf{X}}^0$ specificata come una Bayesian Network \mathcal{B} su \mathbf{X}

⁹ Con il termine "nodo padre", o *parent node*, si intende un nodo il cui stato condiziona quello di un altro nodo del modello grafico.

- 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, \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 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 la sezione 1.1 a pagina 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 Continuous 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.

1.4 APPENDIMENTO

In questa sezione si argomenta sulla probabilità di un *insieme di dati completo* rispetto a una Continuous time Bayesian Network. A tal fine si mostra come una CTBN 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 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 1.16 (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'istanziamento di tutte le variabili casuali è osservabile per ogni istante temporale di ogni traiettoria).

1.4.1 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'istanziamento 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.

1.4.2 Likelihood

Al fine di presentare il calcolo della likelihood¹⁰ di una CTBN rispetto a un dataset completo $\mathcal{D} = \{\delta_1, \dots, \delta_h\}$ 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 $d = \langle x_d, t_d, x_{d'} \rangle \in \delta$, 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:

$$\begin{aligned} L_X(\mathbf{q}, \boldsymbol{\theta} : d) &= L_X(\mathbf{q} : d) \cdot L_X(\boldsymbol{\theta} : d) \\ &= \mathbf{q}_{x_d} e^{-\mathbf{q}_{x_d} t_d} \cdot \boldsymbol{\theta}_{x_d x_{d'}}. \end{aligned} \quad (1.6)$$

Si noti che l'equazione 1.6 è ricavata moltiplicando la *funzione di distribuzione di probabilità* di $X(t)$ (equazione 1.5) per la “*probabilità istantanea*” di transizione (si veda la definizione 1.7).

¹⁰ La likelihood di un insieme di valori dei parametri, dato un insieme di dati, corrisponde alla probabilità dell'insieme dei dati, dati tali valori dei parametri.

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 sintetizzabile 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:

$$\begin{aligned} L_X(\mathbf{q}, \boldsymbol{\theta} : \mathcal{D}) &= \left[\prod_{d \in \mathcal{D}} L_X(\mathbf{q} : d) \right] \left[\prod_{d \in \mathcal{D}} L_X(\boldsymbol{\theta} : d) \right] \\ &= \left[\prod_x (\mathbf{q}_x)^{M[x]} e^{-\mathbf{q}_x^T [x]} \right] \left[\prod_x \prod_{x' \neq x} (\theta_{xx'})^{M[x, x']} \right]. \end{aligned} \quad (1.7)$$

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 $\text{Pa}(X_i)$ dei suoi nodi genitori. Poiché, nel caso di *dati completi*, si conosce sempre l'istanziamento di $\text{Pa}(X_i)$, allora, per ogni istante di tempo t , si conosce quale matrice di intensità $\mathbf{Q}_{X_i | \text{Pa}_i(x)}$, con $\text{Pa}_i(x) \in \text{Pa}(X_i)$, 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 :

$$\begin{aligned} L_{\mathcal{N}}(\mathbf{q}, \boldsymbol{\theta} : \mathcal{D}) &= \prod_{X_i \in \mathbf{X}} L_{X_i}(\mathbf{q}_{X_i | \text{Pa}(X_i)}, \boldsymbol{\theta}_{X_i | \text{Pa}(X_i)} : \mathcal{D}) \\ &= \prod_{X_i \in \mathbf{X}} L_{X_i}(\mathbf{q}_{X_i | \text{Pa}(X_i)} : \mathcal{D}) L_{X_i}(\boldsymbol{\theta}_{X_i | \text{Pa}(X_i)} : \mathcal{D}). \end{aligned} \quad (1.8)$$

Il termine $L_X(\boldsymbol{\theta}_{X | \text{Pa}(X)} : \mathcal{D})$ esprime la likelihood delle transizioni tra stati. Tale termine trascura il tempo che intercorre fra le transizioni 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:

$$L_X(\boldsymbol{\theta}_{X | \text{Pa}(X)} : \mathcal{D}) = \prod_{\text{Pa}_i(x)} \prod_x \prod_{x' \neq x} (\theta_{xx' | \text{Pa}_i(x)})^{M[x, x' | \text{Pa}_i(x)]}. \quad (1.9)$$

Per quanto riguarda il calcolo di $L_X(\mathbf{q}_{X | \text{Pa}(X)} : \mathcal{D})$ 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 istanziamento per l'insieme dei genitori $\text{Pa}(X)$). 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 $\text{Pa}(X)$ 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 (equazione 1.5):

$$1 - F(t) = e^{-\mathbf{q}_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})}t}.$$

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

$$L_X(\mathbf{q}_{X|\mathbf{p}\mathbf{a}(X)} : \mathcal{D}) = \prod_{\mathbf{p}\mathbf{a}_i(\mathbf{x})} \prod_{\mathbf{x}} (\mathbf{q}_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})})^{M[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})]} e^{-\mathbf{q}_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})} T[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})]}.$$
(1.10)

Combinando l'equazione 1.10 e l'equazione 1.9 si ottiene la likelihood di un dataset completo \mathcal{D} rispetto a un singolo processo di Markov condizionale:

$$L_X(\mathbf{q}, \boldsymbol{\theta} : \mathcal{D}) = \prod_{\mathbf{p}\mathbf{a}_i(\mathbf{x})} \prod_{\mathbf{x}} \left[(\mathbf{q}_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})})^{M[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})]} e^{-\mathbf{q}_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})} T[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})]} \cdot \prod_{\mathbf{x} \neq \mathbf{x}'} (\boldsymbol{\theta}_{\mathbf{x}\mathbf{x}'|\mathbf{p}\mathbf{a}_i(\mathbf{x})})^{M[\mathbf{x}, \mathbf{x}'|\mathbf{p}\mathbf{a}_i(\mathbf{x})]} \right].$$
(1.11)

Si noti che, dal punto di vista algebrico, è conveniente riformulare l'equazione 1.11 come *log-likelihood*:

$$\ell_X(\mathbf{q}, \boldsymbol{\theta} : \mathcal{D}) = \sum_{\mathbf{p}\mathbf{a}_i(\mathbf{x})} \sum_{\mathbf{x}} \left[M[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})] \ln(\mathbf{q}_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})}) - \mathbf{q}_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})} T[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})] + \sum_{\mathbf{x} \neq \mathbf{x}'} M[\mathbf{x}, \mathbf{x}'|\mathbf{p}\mathbf{a}_i(\mathbf{x})] \ln(\boldsymbol{\theta}_{\mathbf{x}\mathbf{x}'|\mathbf{p}\mathbf{a}_i(\mathbf{x})}) \right].$$
(1.12)

È ora possibile asserire che la *log-likelihood* di \mathcal{N} (dall'equazione 1.8) è:

$$\ell_{\mathcal{N}}(\mathbf{q}, \boldsymbol{\theta} : \mathcal{D}) = \sum_{X_i \in \mathbf{X}} \ell_{X_i}(\mathbf{q}, \boldsymbol{\theta} : \mathcal{D}).$$
(1.13)

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.

1.4.3 Stima dei parametri

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

Quando si tratta con dati multinomiali ci sono principalmente due scelte che è possibile fare. La scelta più semplice consiste nell'effettuare una stima dei parametri del modello tramite un approccio *maximum-likelihood*. Tuttavia, è noto che tale approccio può portare a problemi con l'inferenza quando i dati di input sono sparsi. Per evitare tale limitazione solitamente si effettua una *regolarizzazione bayesiana* dei parametri: si sceglie una distribuzione a priori per i parametri e li si aggiorna in accordo ai dati di input.

La stima dei parametri non è un processo fine a se stesso, in quanto, da essi è possibile costruire le matrici di intensità condizionali (CIM) di ogni nodo della CTBN. 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 1.10). Perciò, fissato $pa_i(x)$, si può computare la rispettiva matrice di intensità per un nodo qualsiasi ponendo sulla diagonale il rispettivo vettore dei parametri $q_{x|pa_i(x)}$ e ricavando i valori non sulla diagonale dalla relazione (si veda il teorema 1.2) fra i parametri q e θ :

Come costruire una CIM dai parametri.

$$q_{xx'|pa_i(x)} = \theta_{xx'|pa_i(x)} \cdot q_{x|pa_i(x)}. \quad (1.14)$$

Infine, come vedremo in seguito nel capitolo 3, i parametri sono anche un componente chiave del processo di apprendimento strutturale.

Stima maximum-likelihood

In base a quanto attestato dalla definizione stessa delle CTBN (definizione 1.15 a pagina 12), 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 1.7, 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 la probabilità di transizione uscenti dallo stato x verso x' (sempre fermo restando il condizionamento dato dall'istanziamento dei nodi genitori).

La media della distribuzione esponenziale in questione 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' sapendo che avverrà una transizione è data dal rapporto tra il numero totale di transizioni da x a x' diviso il numero totale di transizioni uscenti da x ; cioè:

$$\frac{M[x, x'|pa_i(x)]}{M[x|pa_i(x)]}.$$

Teorema 1.3. *Parametri maximum-likelihood (MLE). I parametri che massimizzano la likelihood (equazione 1.13) di una Continuous time Bayesian Network sono funzione delle statistiche sufficienti:*

$$\begin{aligned} \mathbf{q}_{\mathbf{x}|\mathbf{pa}_i(\mathbf{x})} &= \frac{M[\mathbf{x}|\mathbf{pa}_i(\mathbf{x})]}{T[\mathbf{x}|\mathbf{pa}_i(\mathbf{x})]} \\ \theta_{\mathbf{x}\mathbf{x}'|\mathbf{pa}_i(\mathbf{x})} &= \frac{M[\mathbf{x}, \mathbf{x}'|\mathbf{pa}_i(\mathbf{x})]}{M[\mathbf{x}|\mathbf{pa}_i(\mathbf{x})]}. \end{aligned} \quad (1.15)$$

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

Stima bayesiana

Un approccio alternativo alla stima dei parametri è la stima bayesiana (si veda Nodelman, 2007, sottosezione 5.1.1).

A tal fine è necessario definire una distribuzione a priori sui parametri di una CTBN. Come si è soliti fare in situazioni di questo tipo, per tale distribuzione si sceglie di usare una *distribuzione a priori coniugata*¹¹ poiché ciò risulta conveniente dal punto di vista algebrico (e quindi computazionale). Infatti, una distribuzione a priori coniugata fornisce un'espressione in forma chiusa per la distribuzione a posteriori (alternativamente potrebbe risultare necessario il calcolo di un integrale numerico).

Si consideri innanzitutto un singolo processo di Markov. Si ricorda (si vedano a tal riguardo le definizioni 1.7 e 1.8 a pagina 7 e a pagina 8) che un processo di Markov ha due insiemi di parametri: θ che parametrizzano una *distribuzione multinomiale* e \mathbf{q} che parametrizzano una *distribuzione esponenziale*.

Una distribuzione a priori coniugata per il parametro \mathbf{q} è la *distribuzione Gamma* $P(\mathbf{q}) = \text{Gamma}(\alpha_{\mathbf{x}}, \tau_{\mathbf{x}})$, dove (si veda Nodelman, 2007):

$$P(\mathbf{q}) = \frac{\tau_{\mathbf{x}}^{\alpha_{\mathbf{x}}+1}}{\Gamma(\alpha_{\mathbf{x}}+1)} \mathbf{q}^{\alpha_{\mathbf{x}}} e^{-\mathbf{q}\tau_{\mathbf{x}}}. \quad (1.16)$$

Invece, avendo assunto l'indipendenza dei parametri e poiché la funzione di densità della distribuzione di probabilità di θ , che è una multinomiale, è positiva, per essa si sceglie come priori coniugata la *distribuzione di Dirichlet* $P(\theta) = \text{Dir}(\alpha_{\mathbf{x}\mathbf{x}_1}, \dots, \alpha_{\mathbf{x}\mathbf{x}_K})$ (si veda Hecker-

¹¹ In teoria della probabilità bayesiana, se le distribuzioni a posteriori $P(\theta|\mathbf{x})$ sono nella stessa famiglia della distribuzione a priori $P(\theta)$, le due distribuzioni sono definite coniugate, e la distribuzione a priori è chiamata *distribuzione a priori coniugata* per la verosimiglianza (*likelihood*). Una distribuzione a priori coniugata è conveniente dal punto di vista algebrico in quanto fornisce una espressione in forma chiusa per la distribuzione a posteriori e perché può fornire delle intuizioni circa il modo con cui la funzione di verosimiglianza aggiorna la distribuzione.

man, 1996; Heckerman *et al.*, 1995), la cui funzione di densità (Steck, Harald and Jaakkola, 2002) è:

$$P(\theta) = \frac{\Gamma(\alpha_x)}{\Gamma(\alpha_{x_{x_1}}) \cdot \dots \cdot \Gamma(\alpha_{x_{x_K}})} \theta_{x_{x_K}}^{\alpha_{x_{x_1}}-1} \cdot \dots \cdot \theta_{x_{x_1}}^{\alpha_{x_{x_K}}-1}. \quad (1.17)$$

Nota 1.4.3.1. Si noti che l'iper-parametro α_x , detto *dimensione equivalente del campione*, è costituito dalla somma dei *conteggi immaginari* $\alpha_{x_{x_1}} + \dots + \alpha_{x_{x_K}}$, chiamati anche *pseudo-conteggi* (Steck, Harald and Jaakkola, 2002). Esso può essere pensato come un fattore che esprime la “forza” della distribuzione a priori, in quanto, più esso aumenta, più le stime dei parametri sono regolarizzate, cioè meno estreme. Chiaramente, quando α_x tende a 0 le stime dei parametri tendono alle stime maximum-likelihood. In letteratura (si veda Steck, Harald and Jaakkola, 2002) questo processo è anche chiamato “smoothing”.

Nota 1.4.3.2. L'iper-parametro τ_x , invece, rappresenta una *quantità di tempo immaginaria* che incorpora la credenza della distribuzione a priori sul parametro della distribuzione esponenziale.

Quindi, se si assume che i parametri sono *stocasticamente indipendenti*, cioè che $P(\theta, \mathbf{q}) = P(\theta) P(\mathbf{q})$, allora le distribuzioni a posteriori (i.e., condizionate sui dati) dei parametri \mathbf{q} e θ sono:

$$\begin{aligned} P(\mathbf{q} | \mathcal{D}) &= \text{Gamma}(\alpha_x + M[x], \tau_x + T[x]) \\ P(\theta | \mathcal{D}) &= \text{Dir}(\alpha_{x_{x_1}} + M[x, x_1], \dots, \alpha_{x_{x_K}} + M[x, x_K]). \end{aligned} \quad (1.18)$$

Al fine di generalizzare quest'idea e ottenere una distribuzione a priori coniugata per un'intera CTBN è necessario che essa soddisfi due assunzioni (comuni per le distribuzioni a priori nelle Bayesian Network, si veda Heckerman (1996)): l'indipendenza globale e locale dei parametri. In base all'*indipendenza globale dei parametri* si può scrivere:

$$P(\mathbf{q}, \theta) = \prod_{X_i \in \mathbf{X}} P(\mathbf{q}_{X_i} | p_{\mathbf{a}}(X_i), \theta_{X_i} | p_{\mathbf{a}}(X_i)). \quad (1.19)$$

Invece, dall'*indipendenza locale dei parametri* consegue che è possibile scrivere:

$$P(\mathbf{q}_X | p_{\mathbf{a}}(X), \theta_X | p_{\mathbf{a}}(X)) = \left[\prod_x \prod_{p_{\mathbf{a}_i}(x)} P(\mathbf{q}_x | p_{\mathbf{a}_i}(x)) \right] \left[\prod_x \prod_{p_{\mathbf{a}_i}(x)} P(\theta_x | p_{\mathbf{a}_i}(x)) \right]. \quad (1.20)$$

Se tale distribuzione a priori soddisfa le assunzioni di indipendenza allora anche la distribuzione a posteriori, essendovi coniugata e perciò appartenente alla stessa famiglia parametrica, le soddisferà. In tal caso è possibile mantenere la distribuzione parametrica in forma chiusa e aggiornarla usando le *statistiche sufficienti*:

- $M[x, x' | p_{\mathbf{a}_i}(x)]$ per il parametro $\theta_{x | p_{\mathbf{a}_i}(x)}$
- $M[x | p_{\mathbf{a}_i}(x)]$ e $T[x | p_{\mathbf{a}_i}(x)]$ per il parametro $\mathbf{q}_{x | p_{\mathbf{a}_i}(x)}$.

Data una distribuzione sui parametri è possibile usarla per predire il prossimo evento, mediando la sua probabilità sull'insieme dei possibili valori dei parametri. Questo tipo di previsione è equivalente all'utilizzo dei *valori attesi* dei parametri, i quali hanno la stessa forma dei parametri maximum-likelihood ma considerano i *conteggi immaginari* degli iper-parametri:

$$\begin{aligned}\hat{\mathbf{q}}_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})} &= \frac{\alpha_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})} + \mathbf{M}[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})]}{\tau_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})} + \mathbf{T}[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})]} \\ \hat{\theta}_{\mathbf{x}\mathbf{x}'|\mathbf{p}\mathbf{a}_i(\mathbf{x})} &= \frac{\alpha_{\mathbf{x}\mathbf{x}'|\mathbf{p}\mathbf{a}_i(\mathbf{x})} + \mathbf{M}[\mathbf{x}, \mathbf{x}'|\mathbf{p}\mathbf{a}_i(\mathbf{x})]}{\alpha_{\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})} + \mathbf{M}[\mathbf{x}|\mathbf{p}\mathbf{a}_i(\mathbf{x})]}.\end{aligned}\tag{1.21}$$

Si osservi che questi parametri sono, teoricamente, validi solo per predire una singola transizione, dopo la quale la distribuzione dei parametri andrebbe aggiornata di conseguenza. Tuttavia, è prassi comune non aggiornare la distribuzione dei parametri utilizzando i succitati *valori attesi* anche per la previsione delle transizioni successive.

2 | CLASSIFICAZIONE

La classificazione è un argomento centrale nei campi di ricerca relativi all'apprendimento automatico (anche detto *machine learning*) e l'analisi dei dati. In generale, essa consiste nel processo di assegnare una *classe* (i.e., un'etichetta) a delle istanze descritte da un insieme di attributi. Si parla di *classificazione supervisionata* quando è necessario indurre un classificatore a partire da un insieme di dati composto da istanze già etichettate e utilizzare tale classificatore per classificare nuove istanze di dati. Nel caso in cui, invece, si voglia individuare dei raggruppamenti intrinseci (i.e., *cluster*) a un insieme di dati composto da istanze non etichettate, e creare in corrispondenza di tali raggruppamenti le classi (incognite), si parla invece di *classificazione non supervisionata*.

In questo capitolo viene quindi introdotta una classe di modelli, che prende il nome di Continuous time Bayesian Network classifier (CTBNC), il cui scopo è la *classificazione supervisionata* di traiettorie multi-variate di variabili discrete a *tempo continuo*. Si descrivono due istanze di tale classe: i classificatori Continuous time Naive Bayes (CTNB) e i classificatori Continuous time tree augmented Naive Bayes (CTTANB).

Mentre nella sezione 2.2 si affronta il processo di *apprendimento* in caso di *dati completi* dei CTBNC, nella sezione 2.3 si presenta un algoritmo di *inferenza esatta* per la classe dei CTBNC.

2.1 MODELLO

Al fine di risolvere il succitato problema della classificazione sono stati proposti numerosi approcci. Ad esempio naive Bayes classifier, un classificatore semplice ma robusto proposto da Duda e Hart (1973); rivelatosi essere uno fra i classificatori più performanti (Langley *et al.*, 1992). Esso apprende dai dati la probabilità condizionale di ogni attributo A_i data la classe C . La classificazione di nuove istanze dei dati è effettuata applicando la *regola di Bayes* al fine di calcolare la probabilità della classe C data l'istanziatura di A_1, \dots, A_N e scegliendo quella con la maggiore probabilità a posteriori. Questo calcolo è reso possibile in modo efficiente grazie ad una forte assunzione: tutti gli attributi A_i sono *condizionalmente indipendenti* (si veda la definizione 1.2) tra di loro data evidenza sulla classe C .

Poiché tale assunzione è chiaramente irrealistica, Friedman *et al.* (1997) ha investigato come migliorare ulteriormente le prestazioni del naive

Bayes classifier evitando assunzioni di indipendenza non giustificate dai dati. A tal fine Friedman *et al.* (1997), generalizzando il naive Bayes classifier, ha proposto una classe di modelli di *classificazione supervisionata*, chiamata Bayesian Network classifier (BNC) (di cui fa parte il Tree Augmented Naive Bayes (TAN) classifier, ad esempio) che ereditano dalla teoria delle Bayesian Network (si rimanda alla definizione 1.1 per maggiori dettagli) una rappresentazione fattorizzata delle distribuzioni di probabilità dei nodi attributo e rappresentano esplicitamente le indipendenze condizionali fra essi.

Seguendo le stesse motivazioni, in Stella e Amer (2012) viene formalizzata una classe di modelli di *classificazione supervisionata*, chiamati Continuous time Bayesian Network classifier (CTBNC), derivata dalle CTBN (si veda definizione 1.15).

Di seguito si definiscono quindi i Continuous time Bayesian Network classifier e due istanze di classificatori appartenenti a tale classe: il Continuous time Naive Bayes classifier (CTNBC) e il Continuous time tree augmented Naive Bayes classifier (CTTANBC).

Un Continuous time Bayesian Network classifier estende una CTBN tramite l'aggiunta di un nodo associato alla variabile classe Y . Si ricorda, dalla definizione 1.15, che una CTBN rappresenta l'evoluzione nel tempo continuo di una variabile di processo X (i. e., insieme composto da N processi di Markov, si veda la definizione 1.11).

Di seguito si dà la definizione di questa nuova classe di modelli di *classificazione supervisionata*.

Definizione 2.1 (Continuous time Bayesian Network classifier). Un Continuous time Bayesian Network classifier (CTBNC) è composto da una coppia $\mathcal{C} = (N, P(Y))$ dove:

- N è una CTBN con nodi attributo X_1, X_2, \dots, X_N
- Y è il nodo classe con valori $\text{val}(Y) = \{y_1, \dots, y_K\}$ e probabilità marginale $P(Y)$.

E inoltre il grafo su N (i. e., il grafo \mathcal{G} , si veda la definizione 1.15) rispetta le seguenti condizioni:

- \mathcal{G} è un grafo connesso¹²
- $\text{Pa}(Y) = \{\}$, i. e., la variabile casuale Y è associata al nodo classe
- il nodo Y è indipendente dal tempo ed è specificato solo ed esclusivamente dalla sua probabilità marginale $P(Y)$.

A supporto della definizione 2.1, la figura 2.1 nella pagina seguente fornisce un'istanza di CTBNC composta dai nodi attributi X_1, X_2, X_3, X_4, X_5 e dal nodo classe Y (nodo radice). Si osservi come tale istanza contenga dei cicli, uno riguardante i nodi X_2, X_4, X_5, X_3 e l'altro riguardante

¹² Il grafo $\mathcal{G} = (V, E)$ è detto *connesso* se $\forall (u, v) \in V$ esiste un cammino che collega u a v .

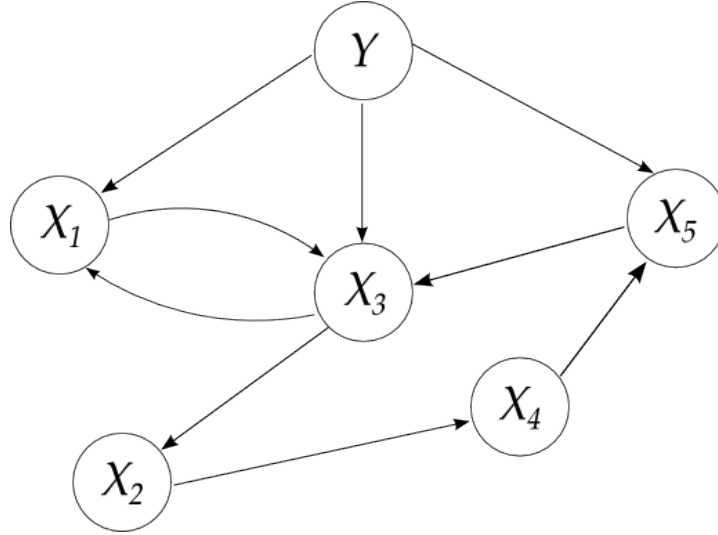


Figura 2.1: Un esempio di Continuous time Bayesian Network classifier (CTBNC) con cinque nodi attributo, X_1, \dots, X_5 , e un nodo classe, Y .

i nodi X_1, X_3 . Si fa notare che gli archi della rete \mathcal{N} rappresentano le dipendenze causali nel tempo.

Parallelamente a quanto fatto in Langley *et al.* (1992), si presentano ora due istanze particolari di Continuous time Bayesian Network classifier.

Definizione 2.2 (Continuous time Naive Bayes classifier). Un Continuous time Naive Bayes classifier (CTNBC) è un Continuous time Bayesian Network classifier $\mathcal{C} = (\mathcal{N}, \mathbf{P}(Y))$ caratterizzato dal fatto che ogni nodo attributo ha un solo genitore, il nodo classe Y . Risulta quindi che:

$$\text{Pa}(X_i) = \{Y\} \quad \forall X_i \in \mathcal{G}.$$

Come mostrato dalla figura 2.2, un CTNBC possiede un nodo radice, associato alla variabile casuale Y , che è l'unico genitore di tutti i restanti nodi X_i (con $i = 1, 2, \dots, N$) che lo compongono. Si osservi come la rete di un CTNBC rappresenti l'assunzione di *indipendenza condizionale* di ogni nodo attributo dagli altri, data evidenza sulla variabile classe Y .

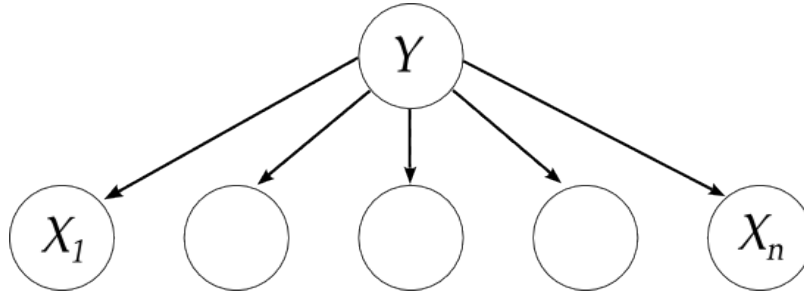


Figura 2.2: Un Continuous time Naive Bayes classifier (CTNBC).

Definizione 2.3 (Continuous time tree augmented Naive Bayes classifier). Un Continuous time tree augmented Naive Bayes classifier (CTTANBC) è un Continuous time Bayesian Network classifier $\mathcal{C} = (\mathcal{N}, \mathbf{P}(Y))$ che rispetta i seguenti vincoli:

- $Y \in \text{Pa}(X_i)$ con $i = 1, 2, \dots, N$
- i nodi attributo X_i , $i = 1, 2, \dots, N$, formano un albero:

$$\exists j : |\text{Pa}(X_j)| = 1 \quad \text{mentre per } i \neq j : |\text{Pa}(X_i)| = 2.$$

Come mostrato dalla figura 2.3, un classificatore CTTANB è un estensione del classificatore CTNB: tutti i nodi attributo della rete \mathcal{N} sono vincolati ad avere come genitore, oltre al nodo radice, al massimo un altro nodo attributo. Ciò comporta che tutti i nodi attributo facciano parte del *Markov blanket* del nodo radice associato con la variabile classe Y .

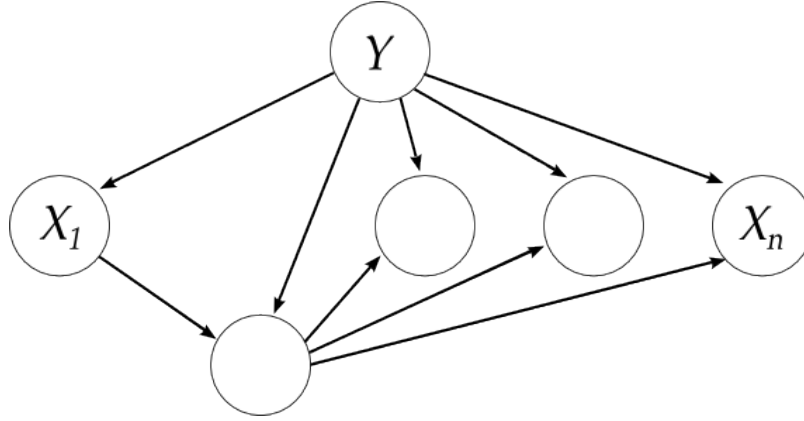


Figura 2.3: Un Continuous time tree augmented Naive Bayes classifier (CTTANBC): qualora la variabile classe Y venga rimossa, le variabili rimanenti formano un albero.

2.2 APPENDIMENTO

In questa sezione si affronta il problema dell'apprendimento (da *dati completi*) dei CTBNC.

Per definizione (si veda la definizione 2.1 a pagina 22) i CTBNC sono basati sul modello delle CTBN, rappresentano perciò un insieme di modelli di probabilità locali (relativi alle variabili casuali) esprimibili in termini di statistiche sufficienti (per maggiori dettagli relativi a questo aspetto si rimanda alla sezione 1.4). Ne deriva che il problema dell'apprendimento di un classificatore CTBN si riduce alla computazione delle statistiche sufficienti dei suoi nodi attributo, da cui è successivamente possibile stimare i parametri (argomento trattato in dettaglio nella sottosezione 1.4.3) delle distribuzioni di probabilità codificate dalle matrici di intensità condizionali (CIM).

Di conseguenza, per l'apprendimento di un CTNBC è richiesto uno sforzo computazionale minimo. Per l'apprendimento di un CTTANBC, poiché questo modello prevede archi anche fra i nodi attributo, è invece richiesto uno sforzo computazionale leggermente maggiore (Stella e Amer, 2012).

Si presenta di seguito l'algoritmo 2.1 (Stella e Amer, 2012) relativo all'apprendimento di un classificatore CTNB (definizione 2.2).

Esso richiede in input un *dataset completo* $\mathcal{D} = \{\delta_1, \dots, \delta_h\}$, il corrispondente insieme delle classi $\{y_1, y_2, \dots, y_h\}$, con $y_i \in \text{val}(Y)$, e il grafo \mathcal{G} di una CTBN \mathcal{N} (rispettivamente chiamati *data*, *classes* e *graph* nella firma della funzione `ctnbclearn`).

Per completezza si osservi (in base alla definizione 1.16) che ogni δ_i (con $i = 1, \dots, h$) è un flusso di evidenze $(x^1, x^2, \dots, x^{j_i})$ (a tal riguardo si veda la definizione 1.14).

Il risultato dell'applicazione dell'algoritmo 2.1 è un Continuous time Naive Bayes classifier $\mathcal{C} = (\mathcal{N}, \mathbf{P}(Y))$.

```

1 function ctnbclearn(data, classes, graph) {
2     var h = len(data)
3     var klass = unique(classes)
4     var nk = len(klass)
5     var priors[nk]
6     for (i in index(data)) {
7         var k = index(classes[i])
8         priors[k] = priors[k] + (1 / h)
9     }
10    var m(), t()
11    for (i in index(data)) {
12        var y = classes[i]
13        var j = 1
14        while ( $t_j \leq T_i$ ) {
15            for (n in index(graph.nodes)) {
16                 $m(x_n^j, x_n^{j+1}, y) = m(x_n^j, x_n^{j+1}, y) + 1$ 
17                 $t(x_n^j, y) = t(x_n^j, y) + (t_j - t_{j-1})$ 
18            }
19            j = j + 1
20        }
21    }
22    var q(), thet()
23    foreach (y in klass) {
24        for (n in index(graph.nodes)) {
25            for ( $x_n$  in  $\text{val}(X_n)$ ) {
26                 $\text{mm}(x_n, y) = \sum_{x'_n \neq x_n} m(x_n, x'_n, y)$ 
27                 $q(x_n, y) = \text{mm}(x_n, y) / t(x_n, y)$ 
28                 $\text{thet}(x_n, y) = m(x_n, x'_n, y) / \text{mm}(x_n, y)$ 
29            }

```

```

30     }
31 }
32 var ctbn = new ctbn(graph, q, thet)
33 return (priors, ctbn)
34 }

```

Algoritmo 2.1: Apprendimento di un classificatore CTNB

L'algoritmo di apprendimento appena presentato consiste nella stima delle matrici di intensità condizionali (CIM) di ogni variabile casuale di \mathcal{N} per ogni classe $y_i \in Y$. Più in dettaglio, esso è composto da tre fasi consecutive:

1. da linea 2 a linea 9 viene calcolata la *probabilità a priori* della variabile classe Y in base alla frequenza di ogni sua istanziazione $y_i \in Y$ in \mathcal{D}
2. da linea 10 a linea 21 vengono calcolate le statistiche sufficienti di ogni nodo attributo X_i , con $i = 1, 2, \dots, N$, sull'insieme di dati di apprendimento (anche detto *training set*) \mathcal{D}
3. da linea 22 a linea 31 vengono infine stimati, a partire dalle statistiche sufficienti, i *parametri maximum-likelihood* (MLE).

Si osservi che, poiché il processo di apprendimento è eseguito su un classificatore CTNB, l'algoritmo condiziona sia il calcolo delle statistiche sufficienti (i.e., variabili t e m) che la stima dei parametri (i.e., variabili q e thet) di ogni nodo attributo X_i solo ed esclusivamente al valore della variabile classe Y (i.e., variabile y). Questa semplificazione è dovuta al vincolo che caratterizza i classificatori CTNB: ogni nodo attributo X_i ha un solo genitore, il nodo associato alla variabile classe Y (si veda la definizione 2.2).

Come anticipato, al costo di un leggero incremento di complessità computazionale, è possibile estendere l'algoritmo 2.1 al fine di creare un algoritmo di apprendimento generale che apprenda un qualsiasi classificatore CTBN. Affinché tale obiettivo sia raggiunto è necessario rimuovere il succitato vincolo sull'insieme dei genitori di ogni nodo attributo. Mentre il calcolo della probabilità a priori della variabile classe Y non varia, il calcolo delle statistiche sufficienti e la stima dei parametri, invece, necessitano di tale generalizzazione.

Nello specifico:

1. il calcolo delle statistiche sufficienti di ogni nodo attributo X_i va condizionato all'istanziatura attuale (i.e., al tempo j) del suo insieme di nodi genitori $\text{Pa}(X_i)$; perciò a linea 15 dell'algoritmo 2.2 si prende in considerazione tale valore (i.e., variabile p)
2. la stima dei parametri maximum-likelihood (MLE) di ogni nodo attributo X_i va eseguita in base a ogni istanziazione del suo insieme di nodi genitori $\text{Pa}(X_i)$; perciò a linea 25 si itera in

base a ogni valore (i.e., variabile p) assunto da $\text{Pa}(X_i)$ (i.e., $\text{val}(\text{Pa}(X_i))$) mentre l'iterazione per classe non è più coerente e di conseguenza rimossa.

Si riporta di seguito l'algoritmo 2.2, il quale include le succitate modifiche finalizzate alla creazione di un algoritmo di apprendimento generale per i CTBNC.

```

1  function learn(data, classes, graph) {
2      var h = len(data)
3      var klass = unique(classes)
4      var nk = len(klass)
5      var priors[nk]
6      for (i in index(data)) {
7          var k = index(klass[i])
8          priors[k] = priors[k] + (1 / h)
9      }
10     var m(), t()
11     foreach (i in index(data)) {
12         var j = 1
13         while ( $t_j \leq T_i$ ) {
14             foreach (n in index(graph.nodes)) {
15                 var p =  $\text{val}^j(\text{Pa}(X_n))$ 
16                  $m(x_n^j, x_n^{j+1}, p) = m(x_n^j, x_n^{j+1}, p) + 1$ 
17                  $t(x_n^j, p) = t(x_n^j, p) + (t_j - t_{j-1})$ 
18             }
19             j = j + 1
20         }
21     }
22     var q(), thet()
23     foreach (n in index(graph.nodes)) {
24         for ( $x_n$  in  $\text{val}(X_n)$ ) {
25             foreach (p in  $\text{val}(\text{Pa}(X_n))$ ) {
26                  $\text{mm}(x_n, p) = \sum_{x'_n \neq x_n} m(x_n, x'_n, p)$ 
27                  $q(x_n, p) = \text{mm}(x_n, p) / t(x_n, p)$ 
28                  $\text{thet}(x_n, p) = m(x_n, x'_n, p) / \text{mm}(x_n, p)$ 
29             }
30         }
31     }
32     var ctbn = new ctbn(graph, q, thet)
33     return (priors, ctbn)
34 }

```

Algoritmo 2.2: Apprendimento di un classificatore CTBN

2.3 INFERENZA

In questa sezione si affronta il problema della *classificazione* di un *flusso di evidenze completamente osservato*, indicato con $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)$ (si veda la definizione 1.14), rispetto a un classificatore CTBN (CTBNC). Si osservi che, in tale situazione (i. e., *dati completi*), l'unica variabile casuale non osservata è la variabile classe, perciò è possibile sfruttare le relazioni di indipendenza fra variabili casuali così come si fa per le Bayesian Network. L'argomento di questa sezione è quindi il processo di *classificazione supervisionata*, di cui si presentano in primis le basi teoriche e successivamente l'implementazione algoritmica che ne consegue.

Il processo di classificazione di un flusso di evidenze è effettuato in base alla regola *maximum a posteriori*¹³ (MAP) (si veda Stella e Amer, 2012): un flusso di evidenze completamente osservato viene classificato assegnandogli la classe la cui probabilità a posteriori (rispetto al flusso di evidenze stesso) è massima. A tale scopo è necessario calcolare la probabilità a posteriori della variabile classe Y del CTBNC, rispetto al flusso di evidenze in input, per tutti i suoi possibili stati (i. e., classi, o etichette).

Il classificatore CTBN classifica quindi il flusso di evidenze massimizzando la seguente probabilità a posteriori, ricavata applicando la *regola di Bayes*:

$$P(Y | (\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)) = \frac{P((\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J) | Y) P(Y)}{P((\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J))}. \quad (2.1)$$

Si specifica di seguito la semantica dei componenti dell'equazione 2.1:

- la probabilità marginale associata alla variabile classe Y

$$P(Y)$$

- la probabilità del flusso di evidenze

$$P((\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J))$$

- la likelihood del flusso di evidenze dato il valore della variabile classe, a cui ci si riferisce nel prosieguo usando l'espressione "*likelihood temporale*"

$$P((\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J) | Y).$$

¹³ La stima della probabilità maximum a posteriori (MAP) è una moda della distribuzione a posteriori che può essere usata per ottenere una stima puntuale di una quantità inosservata sulla base di dati empirici. Può essere vista come una regolarizzazione della stima maximum-likelihood (MLE) poiché è strettamente correlata ad essa; vi differisce perché impiega un obiettivo di massimizzazione incrementato che incorpora una distribuzione a priori sopra la quantità che si vuole stimare.

La probabilità del flusso di evidenze (i. e., denominatore dell'equazione 2.1), similmente a quanto accade per i BNC (Friedman *et al.*, 1997), può essere omessa poiché sussiste la relazione di proporzionalità tra il numeratore e la probabilità che si intende calcolare:

$$P(Y | (\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)) \propto P((\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J) | Y) P(Y). \quad (2.2)$$

Il primo termine dell'equazione 2.2, cioè la *likelihood temporale*, è invece fondamentale per la classificazione tramite regola MAP ed è possibile riformularlo nel seguente modo:

$$P((\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J) | Y) = \prod_{j=1}^J P(\mathbf{x}^j | Y) P(\mathbf{x}^{j+1} | \mathbf{x}^j, Y), \quad (2.3)$$

dove:

- $P(\mathbf{x}^j | Y)$ rappresenta la probabilità che il *vettore aleatorio*¹⁴ \mathbf{X} resti nello stato \mathbf{x}^j durante l'intervallo temporale $[t_{j-1}, t_j)$ data evidenza sulla variabile classe Y
- $P(\mathbf{x}^{j+1} | \mathbf{x}^j, Y)$ rappresenta la probabilità che in \mathbf{X} si verifichi una transizione da \mathbf{x}^j a \mathbf{x}^{j+1} all'istante di tempo t_j data evidenza sulla variabile classe Y .

Inoltre, al fine di assicurare la consistenza dell'equazione 2.3 si assume che $P(\mathbf{x}^{J+1} | \mathbf{x}^J, Y) = 1$.

Il passo successivo consiste nel calcolo dei due termini da cui è composta l'equazione 2.3. A tal fine si utilizzano le distribuzioni di probabilità locali associate ad ogni nodo del CTBN su cui è costruito il classificatore. Come già descritto nella sottosezione 1.1.2, tali modelli di probabilità sono espressi tramite le matrici di intensità condizionali e quindi tramite i parametri \mathbf{q} e $\boldsymbol{\theta}$.

In tale contesto è quindi possibile calcolare il termine $P(\mathbf{x}^j | Y)$ come segue:

$$P(\mathbf{x}^j | Y) = \prod_{n=1}^N \exp\left(-\mathbf{q}_{\mathbf{x}_n^j}^{\mathbf{p}a_j(\mathbf{x}_n)} (t_j - t_{j-1})\right), \quad (2.4)$$

dove $\mathbf{q}_{\mathbf{x}_n^j}^{\mathbf{p}a_j(\mathbf{x}_n)}$ è il valore del parametro della *distribuzione esponenziale* quando la variabile casuale X_n è nello stato x_n durante il j -esimo intervallo temporale $[t_{j-1}, t_j)$ e contemporaneamente l'istanziamento dei genitori di X_n è $\mathbf{p}a_j(\mathbf{x}_n)$.

Uguualmente, il termine $P(\mathbf{x}^{j+1} | \mathbf{x}^j, Y)$ è così calcolabile:

$$P(\mathbf{x}^{j+1} | \mathbf{x}^j, Y) = \prod_{n=1}^N P(x_n^{j+1} | \mathbf{x}_n^j, Y), \quad (2.5)$$

¹⁴ Un *vettore aleatorio* $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è una *n-upla* (composta da n variabili casuali) i cui elementi sono dati da numeri aleatori.

dove

$$P(x_n^{j+1} | x_n^j, Y) = \begin{cases} q_{x_n^j x_n^{j+1}}^{p_{a_j}(x)} & \text{se } x_n^j \neq x_n^{j+1} \\ 1, & \text{altrimenti} \end{cases}. \quad (2.6)$$

Il termine $P(x_n^{j+1} | x_n^j, Y)$ rappresenta la probabilità che nel vettore aleatorio \mathbf{X} si verifichi una transizione dallo stato x^j allo x^{j+1} , dato il valore della variabile classe Y .

Poiché il modello CTBN implica che, ad ogni istante t_j , solo un componente X_n del vettore aleatorio \mathbf{X} può essere soggetto a transizione, allora $P(x_n^{j+1} | x_n^j, Y)$ rappresenta la probabilità che la variabile casuale X_n effettui una transizione da x_n^j a x_n^{j+1} mentre tutti gli altri componenti del vettore aleatorio \mathbf{X} (i.e., X_i con $i \neq n$) non cambiano il proprio stato.

Perciò, come specificato dall'equazione 2.6, nel caso in cui avvenga un cambio di stato in X_n , il termine $P(x_n^{j+1} | x_n^j, Y)$ equivale alla quantità $q_{x_n^j x_n^{j+1}}^{p_{a_j}(x)}$, ricavata dalla relazione fra i parametri (si veda il teorema 1.2). Tale quantità rappresenta il parametro associato alla transizione da x_n^j , stato in cui la variabile casuale X_n si trovava durante il j -esimo intervallo temporale $[t_{j-1}, t_j]$, a x_n^{j+1} , stato in cui X_n si troverà durante il successivo intervallo temporale $[t_j, t_{j+1}]$; data l'istanziatura $p_{a_j}(x_n)$ dei genitori di X_n durante il j -esimo intervallo temporale.

Combinando l'equazione 2.4 e l'equazione 2.5 si ottiene:

$$P(\mathbf{x}^j | Y) P(\mathbf{x}^{j+1} | \mathbf{x}^j, Y) = \prod_{n=1}^N \exp\left(-q_{x_n^j}^{p_{a_j}(x_n)} (t_j - t_{j-1})\right) P(x_n^{j+1} | x_n^j, Y). \quad (2.7)$$

Utilizzando l'equazione 2.7 appena ricavata è possibile riformulare la *likelihood temporale* (equazione 2.3) nel seguente modo:

$$P((\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J) | Y) = \prod_{j=1}^J \prod_{n=1}^N \exp\left(-q_{x_n^j}^{p_{a_j}(x_n)} (t_j - t_{j-1})\right) P(x_n^{j+1} | x_n^j, Y). \quad (2.8)$$

Sostituendo infine l'equazione 2.8 nell'equazione 2.2 si formula definitivamente la probabilità a posteriori della variabile classe Y dato un flusso di evidenze:

$$P(Y | (\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)) \propto P(Y) \cdot \prod_{j=1}^J \prod_{n=1}^N \left[\exp\left(-q_{x_n^j}^{p_{a_j}(x_n)} (t_j - t_{j-1})\right) \cdot P(x_n^{j+1} | x_n^j, Y) \right]. \quad (2.9)$$

Di conseguenza, dato un classificatore CTBN $\mathcal{C} = \{N, P(Y)\}$ e un flusso di evidenze completamente osservato $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)$, la regola MAP seleziona la classe $y^* \in \text{val}(Y)$ massimizzando l'equazione 2.9:

$$y^* = \arg \max_{y \in \text{val}(Y)} P(Y) \prod_{j=1}^J \prod_{n=1}^N \exp\left(-q_{x_n^j}^{p_{a_j}(x_n)} (t_j - t_{j-1})\right) P(x_n^{j+1} | x_n^j, Y) \quad (2.10)$$

Si presenta di seguito l'algoritmo per l'*inferenza esatta* (Stella e Amer, 2012) di un flusso di evidenze completamente osservato rispetto a un classificatore CTBN.

Tuttavia si osservi che tale algoritmo rappresenta le probabilità come *log-probabilità*¹⁵, a causa della succitata convenienza algebrica che ne deriva. Ciò significa che l'algoritmo 2.3 implementa la probabilità a posteriori della classe dato un flusso di evidenze (equazione 2.9) come segue:

$$\ell_{P(Y|\dots)} = \log(P(Y)) + \sum_{j=1}^J \sum_{n=1}^N -q_{x_n^j}^{p_{\alpha_j}(x_n)} (t_j - t_{j-1}) + \log(P(x_n^{j+1} | x_n^j, Y)). \quad (2.11)$$

A tal proposito si noti che, nel caso in cui non avvenga alcun cambio di stato durante un determinato istante di tempo t_j , la quantità $P(x_n^{j+1} | x_n^j, Y)$ dell'equazione 2.11 sarà pari a 1 (si veda l'equazione 2.6), il cui logaritmo è pari a 0. Ciò si riflette nella struttura di controllo condizionale alla linea 11.

```

1 function infer(ctbnc, timeseg, stream) {
2   var priors = ctbnc.priors
3   var logp[len(priors)]
4   for (k in index(priors)) {
5     logp[k] = log(priors[k])
6   }
7   for (k in index(priors)) {
8     for (j in index(timeseg)) {
9       for (n in index(ctbnc.graph.nodes)) {
10        logp[k] = logp[k] -  $q_{x_n^j}^{p_{\alpha_j}(x_n)}$  * timeseg[j]
11        if ( $x_{n_j} \neq x_{n_{j+1}}$ ) {
12          logp[k] = logp[k] + log( $q_{x_n^j x_n^{j+1}}^{p_{\alpha_j}(x)}$ )
13        }
14      }
15    }
16  }
17  return which(max(logp))
18 }
```

Algoritmo 2.3: Inferenza su un classificatore CTBN

L'algoritmo 2.3 di inferenza esatta appena presentato è composto principalmente da due fasi corrispondenti ai due termini principali

¹⁵ La *log-probabilità* è un modalità di rappresentazione della probabilità che porta con sé alcuni vantaggi algebrici e computazionali. Ad esempio, l'utilizzo della *log-probabilità* generalmente comporta una maggior velocità dovuta alla trasformazione delle moltiplicazioni, più computazionalmente costose, in addizioni. In informatica è molto comune l'utilizzo della sua variante negativa, la quale codifica un valore di probabilità $x \in [0, 1]$ come $x' = -\log(x) \in \mathbb{R}$.

dell'equazione 2.11: da linea 2 a linea 6 converte la distribuzione della variabile classe Y del classificatore CTBN in forma logaritmica; da linea 7 a linea 16 aggiorna, incrementandola o decrementandola, il valore di *log-probabilità* relativo a ogni classe (si veda il ciclo **for** alla linea 7) del classificatore CTBN iterando il flusso di evidenze per ogni segmento temporale cui sono associate le sue istanze (i.e., variabile *timeseg*, che si assume venga fornita in input; si veda il ciclo **for** alla linea 8), infine iterando sui nodi del grafo \mathcal{N} del CTBNC (ciclo **for** a linea 9).

Nota 2.3.1. Si osservi che è possibile implementare l'algoritmo di inferenza anche utilizzando la *parametrizzazione mista* (si veda la definizione 1.8). In tal caso si evita la computazione delle matrici di intensità condizionali (CIM) ma, in contrasto, è necessario calcolare il termine $q_{x_n^j, x_{n+1}^i}^{p_{\alpha_j}(x)}$ (a tal proposito si veda l'equazione 1.14) ogni qual volta una variabile casuale del flusso di evidenze di input effettui una transizione (a patto che tale transizione avvenga fra due stati appartenenti allo spazio degli stati della rispettiva variabile casuale associata al nodo del classificatore CTBN).

Il passo finale dell'algoritmo, corrispondente alla linea 17, consiste nella restituzione della classe la cui *log-probabilità* è maggiore di quella delle restanti classi. Tale passo completa perciò l'implementazione dell'equazione 2.10.

L'algoritmo di inferenza esatta presentato in questa sezione è polinomiale.

3 | APPRENDIMENTO STRUTTURALE

Uno dei casi principali che costituisce il problema dell'*apprendimento* di modelli grafico probabilistici è l'apprendimento della struttura incognita sottostante un modello, cioè la selezione di un modello probabilistico che rappresenti un dato *training set*.

Il problema dell'*apprendimento strutturale* da *dati completi* di una Continuous time Bayesian Network (CTBN) è l'argomento trattato in questo capitolo.

Questo problema può essere informalmente descritto nel seguente modo: dato un *training set* composto da istanze di un insieme di variabili casuali si trovi un grafo che rappresenti le relazioni fra le variabili casuali evidenziate nei dati.

L'obiettivo è quindi indurre una struttura (i. e., grafo) che descriva nel miglior modo possibile la distribuzione di probabilità sui dati (i. e., *training set*). Si osservi, inoltre, che questo problema di ottimizzazione è *NP-completo* per le Bayesian Network (Chickering, 1994; Chickering *et al.*, 2004). Per questa ragione viene spesso trattato con algoritmi approssimati.

Per quanto riguarda invece il caso delle CTBN, Nodelman *et al.* (2002) hanno dimostrato che, grazie alla mancanza del vincolo di aciclicità, come già accennato nella sezione 1.3, il problema dell'apprendimento strutturale di una CTBN è significativamente più facile rispetto all'apprendimento strutturale di una Bayesian Network, o di modelli da esse derivanti (e. g., le Dynamic Bayesian Networks (DBN)). Inoltre, nel caso si vincoli la procedura di ricerca a strutture con un numero massimo di genitori per nodo, questo problema può essere risolto in tempo polinomiale rispetto al numero di nodi nella rete¹⁶

L'approccio che si presenta in questo capitolo è quindi un approccio basato sul punteggio: si definisce una funzione che computa uno *score bayesiano* finalizzato alla valutazione di ogni struttura rispetto ai dati di addestramento (i. e., *training set*) e si usa una tecnica di ricerca euristica (e. g., la ricerca *hill climbing*) per cercare nello spazio delle strutture candidate quella che esibisce il maggior punteggio.

Si osservi che l'apprendimento dei parametri (si veda la sottosezione 1.4.3) è propedeutico per tale obiettivo poiché essi costituiscono la base dello score bayesiano.

¹⁶ Si noti comunque che il problema rimane esponenziale rispetto al numero massimo di genitori nella rete. Per tale motivo Stella e Amer (2012) definiscono un classificatore bayesiano a tempo continuo.

3.1 FUNZIONE DI SCORING

Qualsiasi processo di apprendimento strutturale basato su punteggio è costituito da due componenti: una *funzione di scoring* e una procedura di ottimizzazione.

L'obiettivo di questa sezione è quindi presentare una funzione di scoring per l'apprendimento strutturale delle Continuous time Bayesian Network (CTBN). Lo scopo di tale funzione è calcolare il punteggio (i.e., lo score bayesiano) di una struttura relativamente al *training set* \mathcal{D} fornito.

Si definisce lo *score bayesiano* sul grafo \mathcal{G} di una CTBN nel seguente modo:

$$\text{score}_B(\mathcal{G} : \mathcal{D}) = \ln P(\mathcal{D} | \mathcal{G}) + \ln P(\mathcal{G}) \quad (3.1)$$

Come mostra l'equazione 3.1 la funzione di scoring utilizza la probabilità a posteriori dell'insieme dei dati di apprendimento (i.e., il training set \mathcal{D}) data la struttura candidata (i.e., \mathcal{G}), oltre alla probabilità a priori della struttura stessa.

È possibile aumentare in modo significativo l'efficienza dell'algoritmo di ricerca che si affronta nella prossima sezione qualora si facciano determinate assunzioni. Nello specifico, se si assume che la probabilità a priori della struttura, $P(\mathcal{G})$, soddisfi la *structure modularity*, ne consegue:

$$P(\mathcal{G}) = \prod_{X_i} P(\text{Pa}(X_i) = \text{Pa}_{\mathcal{G}}(X_i)). \quad (3.2)$$

Se si assume, inoltre, che la probabilità a priori dei parametri soddisfi la *parameter modularity*, allora per ogni due strutture \mathcal{G} e \mathcal{G}' tali che $\text{Pa}_{\mathcal{G}}(X) = \text{Pa}_{\mathcal{G}'}(X)$ risulta:

$$P(\mathbf{q}_X, \boldsymbol{\theta}_X | \mathcal{G}) = P(\mathbf{q}_X, \boldsymbol{\theta}_X | \mathcal{G}'). \quad (3.3)$$

Combinando l'assunzione di *parameter independence* con l'equazione 3.3 derivante dalla *parameter modularity*, si ottiene:

$$P(\mathbf{q}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{G}} | \mathcal{G}) = \prod_{X_i} \left[P(\mathbf{q}_{X_i} | \text{Pa}(X_i) = \text{Pa}_{\mathcal{G}}(X_i)) \cdot P(\boldsymbol{\theta}_{X_i} | \text{Pa}(X_i) = \text{Pa}_{\mathcal{G}}(X_i)) \right]. \quad (3.4)$$

Si osservi che, poiché la *penalità del grafo*, corrispondente al termine $P(\text{Pa}(X_i) = \text{Pa}_{\mathcal{G}}(X_i))$ dell'equazione 3.2, è legata alla dimensione del grafo ma indipendente dalla quantità dei dati, è possibile ignorare il termine $P(\mathcal{G})$ della funzione di scoring (equazione 3.1).

Di conseguenza il termine significativo dell'equazione 3.1 è la *likelihood marginale*, $P(\mathcal{D} | \mathcal{G})$. Tale termine, infatti, codifica l'incertezza sui parametri integrando su tutti i possibili valori che essi possono assumere:

$$P(\mathcal{D} | \mathcal{G}) = \int_{\mathbf{q}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{G}}} P(\mathcal{D} | \mathbf{q}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{G}}) P(\mathbf{q}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{G}} | \mathcal{G}) d\mathbf{q}_{\mathcal{G}} d\boldsymbol{\theta}_{\mathcal{G}}. \quad (3.5)$$

Come per l'equazione 1.8, la likelihood marginale può essere decomposta come un prodotto di likelihood:

$$\begin{aligned} P(\mathcal{D} | \mathbf{q}_g, \boldsymbol{\theta}_g) &= \prod_{X_i} L_{X_i}(\mathbf{q}_{X_i} | \text{pa}(X_i) : \mathcal{D}) L_{X_i}(\boldsymbol{\theta}_{X_i} | \text{pa}(X_i) : \mathcal{D}) \\ &= \underbrace{\left[\prod_{X_i} L_{X_i}(\mathbf{q}_{X_i} | \text{pa}(X_i) : \mathcal{D}) \right]}_{L(\mathbf{q} : \mathcal{D})} \underbrace{\left[\prod_{X_i} L_{X_i}(\boldsymbol{\theta}_{X_i} | \text{pa}(X_i) : \mathcal{D}) \right]}_{L(\boldsymbol{\theta} : \mathcal{D})}. \end{aligned} \quad (3.6)$$

Combinando tale decomposizione con l'*parameter independence* si può riformulare la likelihood marginale (equazione 3.5) nel seguente modo:

$$\begin{aligned} P(\mathcal{D} | \mathcal{G}) &= \int_{\mathbf{q}_g, \boldsymbol{\theta}_g} L(\mathbf{q}_g : \mathcal{D}) L(\boldsymbol{\theta}_g : \mathcal{D}) P(\mathbf{q}_g) P(\boldsymbol{\theta}_g) d\mathbf{q}_g d\boldsymbol{\theta}_g \\ &= \underbrace{\left[\int_{\mathbf{q}_g} L(\mathbf{q}_g : \mathcal{D}) P(\mathbf{q}_g) d\mathbf{q}_g \right]}_{(a)} \cdot \underbrace{\left[\int_{\boldsymbol{\theta}_g} L(\boldsymbol{\theta}_g : \mathcal{D}) P(\boldsymbol{\theta}_g) d\boldsymbol{\theta}_g \right]}_{(b)}. \end{aligned} \quad (3.7)$$

Ottenuta tale equazione, si affronta di seguito l'analisi e la decomposizione dei due termini che la compongono.

Utilizzando l'assunzione di *local parameter independence*, il termine (a) dell'equazione 3.7 è decomponibile nel seguente modo. Si noti che per brevità si pone $u = \text{pa}_i(x)$.

$$\prod_{X_i} \prod_u \prod_x \int_0^\infty P(\mathbf{q}_{x|u}) \cdot L_{X_i}(\mathbf{q}_{x|u} : \mathcal{D}) d\mathbf{q}_{x|u}. \quad (a)$$

Sostituendo a tale termine la distribuzione a priori coniugata su \mathbf{q} (si veda l'equazione 1.16) e la likelihood delle quantità di tempo trascorse in ogni stato (si veda l'equazione 1.10) si ottiene:

$$\begin{aligned} \prod_{X_i} \prod_u \prod_x \int_0^\infty &\left(\frac{(\tau_{x|u})^{\alpha_{x|u}+1}}{\Gamma(\alpha_{x|u}+1)} (\mathbf{q}_{x|u})^{\alpha_{x|u}} e^{-\mathbf{q}_{x|u} \tau_{x|u}} \cdot \right. \\ &\left. \cdot (\mathbf{q}_{x|u})^{M[x|u]} e^{-\mathbf{q}_{x|u} T[x|u]} \right) d\mathbf{q}_{x|u}. \end{aligned} \quad (a)$$

Si procede semplificando:

$$\prod_{X_i} \prod_u \prod_x \int_0^\infty \frac{(\tau_{x|u})^{\alpha_{x|u}+1} \cdot (\mathbf{q}_{x|u})^{\alpha_{x|u}+M[x|u]}}{\Gamma(\alpha_{x|u}+1) \cdot e^{\mathbf{q}_{x|u}(\tau_{x|u}+T[x|u])}} d\mathbf{q}_{x|u}. \quad (a)$$

E infine, risolvendo l'integrale, si ottiene:

$$\prod_{X_i} \prod_u \prod_x \underbrace{\frac{\Gamma(\alpha_{x|u}+M[x|u]+1)(\tau_{x|u})^{\alpha_{x|u}+1}}{\Gamma(\alpha_{x|u}+1)(\tau_{x|u}+T[x|u])^{\alpha_{x|u}+M[x|u]+1}}}_{\text{MargL}^q(X_i, \text{pa}_g(X_i) : \mathcal{D})}. \quad (a)$$

Relativamente all'analisi del termine (b) dell'equazione 3.7 si osservi che, poiché le distribuzioni sui parametri θ sono di *Dirichlet*, tale operazione è analoga a quella comune per le Bayesian Network.

Ne consegue che il termine (b) si semplifica:

$$\underbrace{\prod_{X_i} \prod_u \prod_x \frac{\Gamma(\alpha_{x|u})}{\Gamma(\alpha_{x|u} + M[x|u])} \cdot \prod_{x \neq x'} \frac{\Gamma(\alpha_{xx'|u} + M[x, x'|u])}{\Gamma(\alpha_{xx'|u})}}_{\text{MargL}^\theta(X_i, \text{Pa}_g(X_i) : \mathcal{D})} \cdot (b)$$

Quindi si può riformulare la likelihood marginale:

$$P(\mathcal{D} | \mathcal{G}) = \prod_{X_i} \text{MargL}^q(X_i, \text{Pa}_g(X_i) : \mathcal{D}) \cdot \text{MargL}^\theta(X_i, \text{Pa}_g(X_i) : \mathcal{D}). \quad (3.8)$$

Al fine di derivare la probabilità a priori della struttura (equazione 3.2) si è già assunto in precedenza che l'ipotesi di *structure modularity* sussista. Perciò, sfruttando tale assunzione e combinando l'equazione 3.2 della probabilità a priori della struttura con la likelihood marginale (equazione 3.8), si ottiene:

$$\begin{aligned} \text{score}_B(\mathcal{G} : \mathcal{D}) &= \sum_{X_i} \left[\ln P(\text{Pa}(X_i) = \text{Pa}_g(X_i)) + \right. \\ &\quad \left. + \ln \text{MargL}^q(X_i, \text{Pa}_g(X_i) : \mathcal{D}) + \right. \\ &\quad \left. + \ln \text{MargL}^\theta(X_i, \text{Pa}_g(X_i) : \mathcal{D}) \right] = \\ &= \sum_{X_i} \text{famscore}_B(X_i, \text{Pa}_g(X_i) : \mathcal{D}). \end{aligned} \quad (3.9)$$

Si è quindi definita la funzione di scoring come una somma di score bayesiani, $\text{famscore}_B(X_i, \text{Pa}_g(X_i) : \mathcal{D})$, relativi ai nodi del grafo \mathcal{G} . Ognuno di tali score bayesiani misura la qualità di $\text{Pa}_g(X_i)$ come insieme dei nodi genitori di X_i , dato l'insieme dei dati di apprendimento \mathcal{D} .

3.2 RICERCA DELLA STRUTTURA

In questa sezione si affronta il secondo passo del processo di apprendimento strutturale: l'utilizzo di una *procedura di ottimizzazione* finalizzata alla ricerca di una struttura che massimizzi lo *score bayesiano*.

Chickering (1994) ha mostrato come il problema di apprendere la struttura ottimale di una Bayesian Network, detto problema *k-learn*, dove k è il numero massimo di genitori per ogni variabile casuale, sia un problema *NP-completo* anche qualora si imponga $k = 2$. La ragione di tale complessità è dovuta al vincolo di aciclicità delle BN (i.e., il grafo di una BN, come da definizione 1.1, deve essere un DAG):

non è perciò possibile determinare l'insieme ottimale dei genitori di ogni nodo di una BN individualmente; poiché la scelta di un insieme di genitori per un nodo restringe la possibilità di scelta relativa ai nodi restanti.

Come già accennato ed intuibile dalla composizione dello score bayesiano (si veda la funzione $\text{famscore}_{\mathcal{B}}$, equazione 3.9), la ricerca della struttura ottimale di un modello Continuous time Bayesian Network (CTBN) è invece notevolmente più semplice rispetto a quello relativo alle BN (o alle DBN). La motivazione di tale vantaggio risiede nel fatto che, poiché gli archi fra i nodi del grado 3 di una CTBN rappresentano l'effetto nel tempo del valore attuale della variabile casuale padre sul valore futuro della variabile casuale figlia, non esiste un vincolo di aciclicità ed è di conseguenza possibile ottimizzare l'insieme dei nodi genitori di un qualsiasi nodo separatamente dagli altri.

Inoltre, qualora si restringa il massimo numero di genitori a un valore k , per ogni variabile casuale X_i di una CTBN, con $i = 1, \dots, N$, si può semplicemente enumerare ogni suo possibile insieme di nodi genitori $\text{Pa}(X_i)$ tale che $|\text{Pa}(X_i)| \leq k$ e calcolarne il rispettivo punteggio $\text{famscore}_{\mathcal{B}}(X_i, \text{Pa}(X_i) : \mathcal{D})$. Infine scegliere come insieme dei nodi genitori di X_i quello con punteggio massimo.

Si definisce perciò il seguente teorema (Nodelman *et al.*, 2002).

Teorema 3.1 (Problema k -learn). *Il problema k -learn per le Continuous time Bayesian Network, fissato k , può essere risolto in tempo polinomiale rispetto al numero di variabili casuali N e alla dimensione dell'insieme di dati \mathcal{D} .*

Si osservi che, fissando k a priori, non è necessario enumerare esaurientemente tutti i possibili insiemi di nodi genitori di ogni nodo di una CTBN. È quindi possibile utilizzare un algoritmo di ricerca euristica di tipo *greedy* per esplorare lo spazio di ricerca. Nella sottosezione 3.2.1 si presenta l'algoritmo scelto per l'apprendimento strutturale della struttura ottimale di un modello CTBN.

3.2.1 Hill Climbing

L'algoritmo *hill climbing* è una tecnica di ricerca euristica finalizzata alla risoluzione di problemi di ottimizzazione i cui stati (i.e., elementi dello spazio di ricerca) contengono tutte le informazioni necessarie a costituire una soluzione (Russell e Norvig, 2003).

L'idea su cui si basa tale algoritmo consiste nell'iniziare la ricerca con una soluzione sub-ottimale e, nei passi successivi, migliorare iterativamente la soluzione, osservando gli stati vicini, finché una qualche condizione di fermata venga raggiunta (e.g., l'algoritmo non può migliorare ulteriormente la soluzione corrente). Il processo di miglioramento è, come già accennato, basato sulla valutazione dello stato corrente tramite una funzione di punteggio.

A differenza di altri algoritmi di ricerca basati sul miglioramento iterativo della soluzione (e.g., *simulated annealing*, *tabu search*), l'algoritmo *hill climbing* si sposta sempre in uno stato che fornisce una soluzione con maggior punteggio (Russell e Norvig, 2003). L'utilizzo di tale algoritmo garantisce il raggiungimento di soluzioni localmente ottime (i.e., soluzioni che non possono essere migliorate considerando solamente la configurazione degli stati vicini) in una quantità di tempo relativamente bassa. Tuttavia esso non garantisce il raggiungimento di una soluzione che costituisca l'ottimo globale, a meno che la funzione rappresentante lo spazio di ricerca non sia convessa. Ciò avviene poiché l'algoritmo smette di effettuare progressi verso la soluzione globalmente ottima nel momento in cui il vicinato della soluzione corrente non permette alcun miglioramento immediato.

Per sorpassare tale limitazione è possibile attuare varie strategie (si veda Russell e Norvig, 2003), quali, ad esempio: l'esplorazione iterativa, a partire da diverse configurazioni iniziali, dello stesso spazio di ricerca (i.e., *random restart hill climbing*); la selezione stocastica del vicinato da esaminare ad ogni passo della ricerca locale, sulla base della probabilità che un dato vicinato porti a un progresso maggiore rispetto ad altri vicini; oppure la scelta di soluzioni non migliorative finalizzata ad una maggiore esplorazione dello spazio di ricerca (i.e., *simulated annealing*).

Nota 3.2.1.1. Questo algoritmo opera una gestione efficiente della memoria poiché non necessita il mantenimento di alcun albero di ricerca: esso conserva solamente l'informazione (i.e., punteggio della soluzione) sullo stato corrente e quello successivo.

L'algoritmo 3.1 illustra la tecnica di ricerca *hill climbing* dato uno spazio degli stati discreto.

```

1 function hillclimbing(problem) {
2   var current_state = start_state(problem)
3   while (true) {
4     var nb = neighbors(current_state)
5     var next_eval = -inf
6     var next_state = null
7     for (x in nb) {
8       var x_score = score(x)
9       if (x_score > next_eval) {
10        next_state = x
11        next_eval = x_score
12      }
13    }
14    if (next_eval ≤ score(current_state)) {
15      break
16    }
17    current_state = next_state

```

```

18     }
19     return current_state
20 }

```

Algoritmo 3.1: Algoritmo *hill climbing* per uno spazio degli stati discreto.

Di seguito si discute l'applicazione di tale algoritmo all'apprendimento strutturale delle CTBN.

Come detto, a causa della mancanza del vincolo di aciclicità, è possibile eseguire la succitata procedura di ottimizzazione in modo indipendente per ogni singolo nodo del modello CTBN in esame. Ciò permette di scomporre il problema dell'apprendimento strutturale in un insieme di problemi della stessa entità, di minore complessità e completamente indipendenti fra di essi; contesto ottimo per un approccio parallelizzato alla risoluzione del problema padre.

Dato uno qualsiasi dei succitati sotto-problemi, lo spazio degli stati che l'algoritmo *hill climbing* può esplorare è composto da tutti i possibili insiemi di genitori con cardinalità minore o uguale a k , il numero massimo di genitori di ogni singolo nodo della CTBN. L'algoritmo *hill climbing* individua l'insieme di genitori ottimale per il nodo in esame, valutando, iterativamente, i possibili insiemi di nodi genitori con cardinalità minore e maggiore di 1 rispetto alla cardinalità dell'insieme di genitori corrente. Si osservi che tale configurazione prevede che l'insieme dei genitori con cardinalità pari a 0 (i.e., insieme vuoto \emptyset) venga anch'esso valutato.

La funzione di *scoring* con cui tali insiemi di genitori vengono valutati (funzione **score**, algoritmo 3.1) corrisponde alla funzione $\text{famscore}_{\mathcal{B}}$, presentata nella sezione 3.1 a pagina 34.

4 | PACKAGE R

Uno degli obiettivi di questo lavoro di tesi è consistito nella creazione di un framework in linguaggio R per le CTBN.

In questo capitolo si descrivono tutti gli strumenti che compongono tale framework.

4.1 R

...

4.2 ANALISI

...

4.3 PACKAGE CTBN

...

4.3.1 Gestione dei dati

...

4.3.2 Apprendimento

...

4.3.3 Inferenza

...

4.3.4 Apprendimento strutturale

...

4.4 CROSS-VALIDAZIONE

...

4.4.1 Metriche di valutazione

...

4.4.2 Package xvalidation

...

Al fine di valutare le prestazioni degli algoritmi di apprendimento e classificazione delle Continuous time Bayesian Network è emersa la necessità di generare dei dataset adeguati a tale scopo.

Come già specificato in precedenza, le CTBN sono un modello stocastico dedito alla rappresentazione dell'evoluzione di sistemi dinamici, cioè di fenomeni che evolvono nel tempo, rappresentati come insiemi di traiettorie multi-variate. Si è quindi scelto di generare dei dataset che rappresentassero un tipico sistema dinamico complesso: il traffico automobilistico su rete urbana.

Tale scelta non è finalizzata al mero esercizio delle tecniche di apprendimento e classificazione descritte e sviluppate, bensì è motivata dalla necessità di individuare delle possibili strategie di ottimizzazione del problema del traffico.

Infatti, il traffico intenso è un problema che affligge tutte le grandi città quotidianamente. La congestione delle reti urbane incide negativamente sulla qualità della vita. Il tempo perso in coda dagli automobilisti ha ripercussioni in primis sulle attività economiche. Alcuni studi a riguardo, risalenti al 2004, stimano le perdite del sistema economico italiano da esso derivanti in 6,4 MLD di € annui (Certet, 2004). A ciò si aggiunga l'impatto negativo del traffico sull'ambiente e quindi sulla salute: l'aumento dei tassi di diossina nell'aria è correlato all'aumento delle malattie respiratorie (e.g., asma) e immunitarie (e.g., allergie), o peggio all'insorgenza di tumori (Gualtieri *et al.*, 2005; Mantecca *et al.*, 2007). Inoltre, elevati livelli di traffico sono fonte di stress da cui possono derivare ulteriori complicazioni. Le problematiche citate evidenziano i costi ingenti che il problema del traffico impone alla collettività. Ne deriva la necessità di affrontare tale problema per cercare di ridurre l'impatto negativo. Tra gli approcci (elencati esaustivamente in Papageorgiou *et al.*, 2005) finalizzati al miglioramento delle condizioni del traffico urbano emerge, per il grande impatto atteso in termini di benefici, l'ottimizzazione dei piani semaforici in base alle condizioni di traffico.

L'ottimizzazione semaforica è un problema complesso: pochi degli approcci proposti in letteratura (Felici *et al.*, 2006; Gershenson, 2008; Park e Messer, 1998; Thorpe e Anderson, 1996) possono essere effettivamente applicati nella realtà, seppur in seguito a grosse semplificazioni dovute alla difficoltà computazionale che li caratterizza. Un modo per risolvere questo problema, non nuovo alla letteratura (si veda Angulo *et al.*, 2011), è semplificare il problema dividendolo in due passi:

- classificazione del profilo di traffico
- selezione del piano semaforico ad esso associato.

In questa tesi affrontiamo il primo passo: la classificazione dei profili di traffico, propedeutico all'ottimizzazione del problema descritto. Inoltre, l'algoritmo di apprendimento strutturale, descritto nel capitolo 3 a pagina 33, fornisce la mappa delle relazioni nel tempo tra i vari sensori della rete stradale; ciò si traduce in uno strumento di osservazione e controllo della stessa.

Con l'ausilio di un software commerciale, Traffic Software Integrated System (TSIS) (versione ≥ 6.2), e di una sua estensione a tempo d'esecuzione appositamente sviluppata al fine di monitorare e tracciare il passaggio dei veicoli, sono stati generati dei dataset contenenti un insieme di documenti rappresentanti la presenza (durante l'evolvere del tempo) di veicoli sui sensori di una rete stradale.

Si precisa che i sensori cui ci si riferisce sono le spire, cioè dei sensori ad induttanza magnetica, posti sotto l'asfalto, che generano un'informazione, detta onda quadra, così composta: stato pari a 1 se un veicolo è presente sul sensore, stato pari a 0 viceversa. La figura 5.1 illustra visivamente quanto appena descritto.

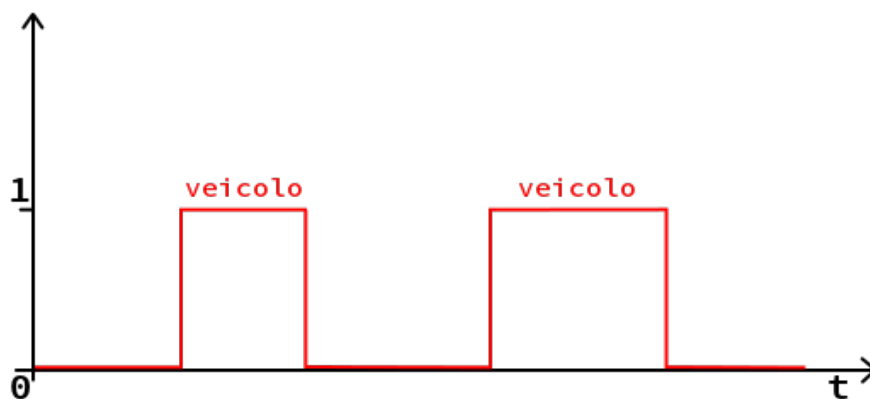


Figura 5.1: Rappresentazione dell'onda quadra generata da un sensore a spira magnetica.

Poiché TSIS supporta esclusivamente la generazione di output aggregati rispetto al tempo (e. g., densità dei veicoli per strada, velocità media, conteggio dei veicoli per sensore), si è reso necessario lo sviluppo di una estensione a tempo d'esecuzione dedicata che fosse in grado di memorizzare il dato atomico dei sensori al fine di ottenere la succitata onda quadra a tempo continuo.

In questo capitolo si presentano sia i succitati strumenti utilizzati per la creazione di reti stradali e relativi modelli di simulazione, sia l'estensione ideata per la creazione di dataset relativi al traffico.

Relativamente, invece, al processo pratico di creazione dei dataset si rimanda alla sezione A.2.

5.1 TSIS

Traffic Software Integrated System (TSIS) è un ambiente di sviluppo integrato¹⁷, distribuito commercialmente da McTrans¹⁸ e supportato dalla Federal Highway Administration (FHWA)¹⁹, il cui scopo ultimo è permettere la simulazione e l'analisi di modelli di reti stradali.

TSIS è costituito da insieme di strumenti dedicati alla creazione di reti stradali e relativi modelli di simulazione, all'esecuzione, e eventualmente alla visualizzazione, di tali modelli, così come all'interpretazione dei risultati ottenuti. Tale insieme di strumenti è reso accessibile tramite delle interfacce grafiche²⁰.

L'architettura modulare con cui TSIS è realizzato permette, in caso di necessità, di estendere tale ambiente creando degli ulteriori strumenti.

Di seguito si introducono brevemente i concetti relativi a TSIS utilizzati nel prosieguo di questo lavoro di tesi.

Tuttavia si osservi che, poiché lo scopo di questa sezione non consiste nel documentare TSIS, la sua trattazione esaustiva (e.g., semantica e lista completa dei tipi di dati rappresentabili) è omessa. A tale fine si rimanda invece alla documentazione ufficiale del software in questione.

Definizione 5.1 (Progetto TSIS). Un progetto TSIS è un insieme di modelli di simulazione per una specifica rete stradale.

Definizione 5.2 (Modello di simulazione TSIS). Un modello di simulazione è costituito da un input (e.g., variazioni dei flussi di ingresso nella rete, variazioni delle percentuali di svolta dei veicoli nelle intersezioni) per la simulazione di una determinata rete stradale e tutti i dati generati dalla sua esecuzione (i.e., simulazione).

Osservazione 5.2.1. Un modello di simulazione può anche prevedere che la sua esecuzione sia eseguita più volte. Finché il seme dei numeri casuali non è modificato, esso è sempre considerato un singolo modello di simulazione.

Definizione 5.3 (Formato TRF). TRF²¹ è il formato dei file accettati dal simulatore di TSIS. Esso codifica e rappresenta una rete stradale e

¹⁷ Un ambiente di sviluppo integrato, comunemente chiamato anche Integrated Development Environment (IDE), è un insieme di programmi finalizzati a supportare il processo di sviluppo dei software. Generalmente, un IDE è costituito da uno strumento per la creazione e modifica del codice sorgente, un compilatore o un interprete, strumenti per l'automazione dello sviluppo e la qualità del codice sorgente.

¹⁸ McTrans Moving Technology: <http://mctrans.ce.ufl.edu>.

¹⁹ Agenzia del Dipartimento dei Trasporti degli Stati Uniti d'America: www.fhwa.dot.gov.

²⁰ Un'interfaccia grafica, nota anche come Graphical User Interface (GUI), è un tipo di interfaccia utente il cui fine è permettere all'utente di interagire con il software manipolando oggetti grafici convenzionali.

²¹ Traffic File (TRF).

il relativo modello di simulazione specificandone i vari componenti tramite l'utilizzo dei rispettivi RT (si veda la definizione 5.5).

Osservazione 5.3.1. Il formato TRF è equivalente al formato TRAF²².

Definizione 5.4 (Formato TNO). TNO²³ è il formato nativo con cui vengono rappresentate in memoria le reti stradali create visivamente tramite l'interfaccia grafica TRAFED.

Definizione 5.5 (Record Type). Un Record Type (RT) rappresenta il blocco informativo minimo su cui è costruita una rete stradale e il suo modello di simulazione. Nel concreto esso consiste in una singola riga di testo nei file TRF contenente un codice identificativo numerico e dei valori per i rispettivi campi accettati nell'ordine prestabilito. Allo stesso modo, tale formato descrive anche il modello di simulazione della rete stradale.

Definizione 5.6 (RTE). Una Run-Time Extension (RTE), estensione a tempo d'esecuzione, è un'applicazione in grado di comunicare a tempo d'esecuzione con un'altra applicazione esterna. Una RTE, in ambiente *Microsoft Windows*, è solitamente compilata separatamente sotto forma di DLL²⁴ e deve rispettare una determinata interfaccia: deve perciò implementare e esportare determinate funzioni che l'applicazione oggetto della comunicazione chiama a tempo d'esecuzione. Per la comunicazione in ingresso, invece, una RTE deve essere collegata alle librerie dell'applicazione con cui intende interfacciarsi, avendo così accesso alle strutture dati e alle funzioni che questa esporta.

5.1.1 Componenti

In questa sezione si elencano gli strumenti che costituiscono l'ambiente di sviluppo TSIS.

CORSIM

CORSIM²⁵ costituisce il componente principale dell'insieme di strumenti denominato TSIS. È un simulatore il cui obiettivo è permettere la creazione e l'esecuzione di modelli di simulazione TSIS. È composto da due simulatori integrati che rappresentano l'intero sistema di traffico come funzione del tempo: NETSIM²⁶

²² Traffic File (TRAF).

²³ TRAFED Native Object (TNO).

²⁴ Una libreria a collegamento dinamico (DLL) è una libreria software che viene caricata dinamicamente in fase di esecuzione, invece di essere collegata staticamente a un eseguibile in fase di compilazione. Queste librerie sono anche chiamate Dynamic-link library (DLL). L'acronimo DLL corrisponde all'estensione che tali oggetti hanno in ambiente *Microsoft Windows*. Tuttavia esse sono spesso chiamate più genericamente con il termine librerie condivise (da *shared library*). Nei sistemi *Linux*, esse sono anche note come oggetti *shared object*.

²⁵ Corridor microscopic simulation program (CORSIM).

²⁶ Network Simulator (NETSIM).

e FRESIM²⁷. Tali simulatori integrati rappresentano, rispettivamente, il traffico sulle strade urbane e non. La simulazione effettuata da tali strumenti è di tipo microscopico: essi modellano individualmente il comportamento di ogni singolo veicolo, prendendo in considerazione per ognuno di essi una serie di variabili, anche di tipo stocastico (e.g., tipologia di guidatore). Per tale motivo CORSIM è dotato di molte possibili opzioni di configurazione e permette lo studio di modelli molto complessi e dettagliati.

TRAFED

TRAFED²⁸ è una GUI il cui scopo è permettere la creazione e la modifica di reti stradali e di modelli di simulazione per CORSIM.

TShell

TShell²⁹ è la GUI di TSIS. Funge da contenitore degli strumenti (preconfigurati, o creati dall'utente) di questo ambiente di sviluppo integrato e permette la gestione dei progetti TSIS.

TRAFVU

TRAFVU³⁰ è una GUI finalizzata alla visualizzazione dei modelli di simulazione simulati con CORSIM. Essa permette sia di visualizzare in modo animato l'evoluzione del traffico nella rete stradale con una qualsiasi granularità temporale, sia di visualizzare una serie di misure di interesse relative alla simulazione.

TSIS TEXT EDITOR

TSIS Text Editor è uno strumento il cui scopo è facilitare la modifica manuale dei file TRF. A tale scopo esso visualizza per ogni RT che si intende modificare sia la sua descrizione sia l'insieme dei campi supportati.

TSIS SCRIPT TOOL

TSIS Script Tool è uno strumento per la creazione, la modifica e l'esecuzione di codice VBScript³¹. Questo strumento fornisce un meccanismo utile ad automatizzare le funzionalità di simulazione di TSIS (e.g., esecuzioni multiple dello stesso modello di simulazione variando il seme dei numeri casuali che governa la distribuzione di ingresso dei veicoli nella rete stradale).

TSIS TRANSLATOR

TSIS Translator è uno strumento utile alla conversione dei file dal

²⁷ Freeway Simulator (FRESIM).

²⁸ TRAF Editor (TRAFED).

²⁹ TSIS Shell (TShell).

³⁰ TRAF Visualization Utility (TRAFVU).

³¹ Microsoft's Visual Basic Scripting Edition (VBScript) è un linguaggio interpretato, sottoinsieme del linguaggio *Visual Basic*, utilizzato come sostituto o integrazione della linea di comando o per il controllo di applicazioni esterne in ambiente *Microsoft Windows*.

formato TRF al formato TNO e viceversa. Tale operazione risulta utile al fine di rendere i file TRF utilizzabili tramite lo strumento TRAFED così come per rendere i file TNO utilizzabili con CORSIM.

TSIS OUTPUT PROCESSOR

TSIS Output Processor è uno strumento finalizzato alla raccolta e l'aggregazione dei dati da CORSIM durante l'esecuzione multipla di modelli di simulazione. La sua caratteristica principale consiste nella computazione automatica di un insieme di statistiche predefinite. Esso permette di scegliere sia le statistiche di interesse sia la granularità temporale della loro computazione.

5.1.2 Caratteristiche

Segue una panoramica il cui scopo è presentare sia le principali capacità, sia i vincoli di modellazione, simulazione e analisi di TSIS.

TSIS, tramite CORSIM, permette la modellazione di reti stradali con le seguenti caratteristiche:

- reti stradali urbane, non urbane o miste
- controllo quantitativo (i. e., quantità di veicoli per intervallo temporale, distribuita secondo una distribuzione statistica) e qualitativo (i. e., tipo di veicoli) dei flussi di veicoli in ingresso nella rete stradale
- controllo completo del flusso di traffico (i. e., percentuali di svolta dei veicoli variabile nel tempo)
- supporto per strade con più corsie (massimo 9)
- supporto dei segnali stradali
- supporto per gli attraversamenti pedonali
- intersezioni non controllate
- intersezioni controllate da semafori
- canalizzazione delle corsie
- piani semaforici predefiniti, dinamici (e.g., priorità ai veicoli di emergenza) o guidati da algoritmi (e.g., attivazione del semaforo in base alle rilevazioni effettuate dai sensori)
- sensori per la rilevazione del passaggio o della presenza di veicoli
- supporto per i mezzi di trasporto pubblico (i. e., autobus, taxi)
- rilevamento di incidenti

- supporto per i sistemi di guida anglosassoni
- simulazione guidata all'analisi delle code
- simulazione guidata allo studio del grado di occupazione della rete stradale
- output di dati di interesse aggregati
- output di dati statistici collezionati

Poiché un modello di simulazione del traffico è caratterizzato dal cambiamento di un insieme di condizioni (e.g., volumi di traffico, canalizzazione delle corsie, percentuali di svolta dei veicoli) della rete stradale, esso deve specificare, oltre alla natura stessa dei cambiamenti, anche i fattori in base a cui essi avvengono. Il fattore primario che viene preso in considerazione è il tempo. Ne consegue perciò che un modello di simulazione del traffico deve specificare l'intervallo temporale in cui specifici cambiamenti di determinate condizioni avvengono.

CORSIM affronta questo problema permettendo di partizionare il tempo totale di simulazione in una serie di periodi temporali (i.e., *time period*) di durata variabile. Ogni periodo temporale possiede perciò un insieme di dati di input che non variano per tutta la sua durata. Inoltre, i periodi temporali sono a loro volta suddivisi in intervalli temporali (i.e., *time interval*), anch'essi suddivisi in passi temporali (i.e., *time step*).

Le principali limitazioni di TSIS derivano dalle modalità con cui CORSIM implementa la gestione del tempo: la lista dei periodi temporali è rappresentata internamente tramite un array statico di dimensione prefissata. Nello specifico, un modello CORSIM può essere costituito da un massimo di 19 periodi temporali, ognuno dei quali viene specificato tramite il RT 03, e può avere una durata compresa tra i 10 e i 9999 secondi. Ognuno di tali periodi temporali è suddiviso in intervalli (specificati tramite dei RT 04) la cui durata, compresa tra 1 e 200 secondi, deve essere un sottomultiplo della durata del periodo temporale cui appartiene. Infine, ogni intervallo temporale è partizionato in passi temporali: NETSIM utilizza un passo temporale fisso di 1 secondo mentre FRESIM opera in base al passo temporale specificato dall'utente. Poiché i due modelli di simulazione microscopica che costituiscono CORSIM operano in modo sincronizzato, la durata del passo temporale di FRESIM, anche se compresa nell'intervallo $[0.1, 1]$ secondi, non può essere specificata liberamente dall'utente. All'utente è permesso specificare solo il numero di passi temporali di FRESIM (tramite il campo 1, RT 04) che devono essere eseguiti per ogni secondo di simulazione. Tale approccio permette perciò di utilizzare dei passi temporali di durata minore per il simulatore FRESIM mantenendo la sincronizzazione con il simulatore NETSIM.

La tabella 5.1 mostra la corrispondenza tra il numero di passi temporali di FRESIM e la reale durata che verrà loro assegnata.

Valore del campo 1	Durata del passo temporale in FRESIM (sec)
1	1.0
2	0.5
3	0.333333
4	0.25
5	0.2
6	0.166667
7	0.142857
8	0.125
9	0.111111
10	0.1

Tabella 5.1: Relazione tra numero di passi temporali per ogni secondo di simulazione e durata effettiva del passo temporale in FRESIM.

La figura 5.2 nella pagina successiva mostra, tramite un esempio, quanto appena descritto relativamente alla rappresentazione del tempo in CORSIM.

Si osservi, inoltre, che la granularità temporale massima con cui è possibile ottenere statistiche cumulative, e, in generale, dati relativi alla simulazione, corrisponde alla durata scelta per gli intervalli temporali (si veda a tal riguardo la documentazione relativa al tipo RT 05). Si consideri ad esempio la figura 5.2 nella pagina seguente: in tal caso non sarà possibile ottenere dati cumulati relativi a un intervallo di tempo minore di 60 secondi.

La gestione del tempo che CORSIM attua implica quindi alcune limitazioni:

- il tempo massimo di simulazione, benché sufficiente nella maggior parte dei casi, è di circa 52 ore
- la massima granularità con cui è possibile raccogliere informazioni è di 1 secondo, impostando a tale valore la durata degli intervalli temporali
- i dati raccolti dalla simulazione sono sempre aggregati in base alla durata dell'intervallo temporale cui si riferiscono.

Tali vincoli rendono impossibile recuperare l'output dei sensori a tempo continuo in modo non aggregato o in generale con una granularità temporale inferiore a 1 secondo. Al fine di sorpassare tali limitazioni si è proceduto sviluppando una estensione a tempo d'esecuzione apposita, *Sensors DLL*, descritta dettagliatamente nella sezione 5.3 a

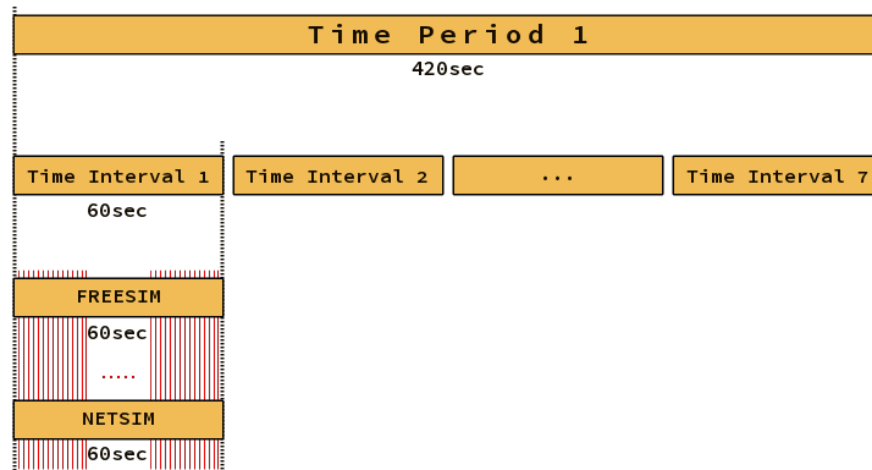


Figura 5.2: Esempio di suddivisione gerarchica delle unità temporali in CORSIM. Un *time period* della durata di 420 secondi è suddiviso in 7 *time interval*, ciascuno della durata di 60 secondi. Ogni *time interval* verrà poi automaticamente diviso in 60 *time step* da 1 secondo ciascuno per NETSIM (i.e., linee rosse). La durata di tali *time step* di NETSIM può o meno coincidere, come accade in questo esempio, con quella assegnata ai *time step* di FRESIM a seconda del valore del campo 1 del RT 04 nel modello CORSIM.

pagina 60. Al fine di rendere la descrizione di tale software maggiormente chiara, è necessario presentare il meccanismo di estensione di CORSIM. La sezione che segue affronta perciò tale argomento.

5.2 CREAZIONE DI ESTENSIONI TSIS

TSIS espone un meccanismo finalizzato all'estensione delle sue funzionalità tramite la creazione, da parte dell'utente, di altri strumenti da integrare nell'ambiente di sviluppo. Tali strumenti, interfacciandosi direttamente con CORSIM, possono modificarne o aumentarne la logica di simulazione, collezionare dati o monitorare eventi speciali (e.g., incidenti).

In questa sottosezione si presenta il funzionamento dei meccanismi di interfacciamento (i.e., più brevemente detti API³²) tra CORSIM e strumenti esterni, a cui ci si riferirà da questo momento in poi con il termine CORSIM RTE.

³² Con il termine Application Programming Interface (API) si indica un insieme di procedure rese disponibili all'esterno, di solito raggruppate a formare un insieme di strumenti specifici per l'espletamento di un determinato compito all'interno di un programma.

5.2.1 Requisiti

Al fine di sviluppare e compilare con successo una CORSIM RTE in C++ è necessario disporre dei seguenti strumenti:

- il compilatore della piattaforma *Microsoft Visual C++*
- il pacchetto software di TSIS, il quale include di default tutti i componenti necessari mostrati dalla figura 5.3 nella pagina seguente (ad eccezione, chiaramente, del componente RTE).

Si osservi, inoltre, che è possibile sviluppare una CORSIM RTE anche in linguaggio C o FORTRAN.

5.2.2 Architettura di CORSIM

La figura 5.3 nella pagina successiva illustra l'architettura modulare di CORSIM e il suo funzionamento all'interno dell'ambiente di sviluppo TSIS. Si osservi che i componenti che costituiscono CORSIM sono di due tipi: librerie DLL e moduli COM³³. Il CORSIM Driver Component, ad esempio, è il modulo COM di TSIS preposto a interfacciare CORSIM e TShell, permettendo così il controllo e l'esecuzione di CORSIM, delle RTE create dall'utente e degli altri strumenti (e. g., TSIS Output Processor) di TSIS tramite GUI.

Anche se la figura 5.3 nella pagina seguente mostra per completezza l'intera architettura di CORSIM, si procede con la descrizione delle interfacce di CORSIM preposte alla comunicazione con RTE sviluppate dall'utente, identificate dalle frecce tratteggiate e numerate.

Per ogni passo temporale di simulazione, il CORSIM Server chiama una serie di funzioni di CORSIM finalizzate a guidare l'andamento della simulazione. Quando una RTE viene inserita nell'ambiente di sviluppo integrato, il CORSIM Server chiama anche le funzioni che la RTE esporta in base ai messaggi che riceve da CORSIM durante la sua esecuzione. Questa interfaccia è rappresentata dalla freccia 1 nella figura 5.3 nella pagina successiva; la sottosezione 5.2.3 a pagina 53 riporta maggiori dettagli sui punti di chiamata che CORSIM espone.

TSIS fornisce inoltre una API, identificata dalla freccia 2 nella figura 5.3 nella pagina successiva, chiamata CORWIN, che permette alle RTE di inviare messaggi al modulo CORSIM Server affinché essi siano visualizzati in TShell dal CORSIM Driver Component.

Infine, una RTE può accedere direttamente a una serie di funzioni e strutture dati esportate da CORSIM nella memoria condivisa. Anche se non è possibile riferirsi a questo meccanismo di comunicazione come una API vera e propria, nel prosieguo, ci riferiremo ad essa con il

³³ Il Component Object Model (COM) è uno standard per componenti software ideato da *Microsoft*. Il suo fine consiste nel permettere la comunicazione fra processi e la creazione dinamica di oggetti. Una interfaccia COM è una collezione di funzioni, incapsulata in un componente software binario e neutrale rispetto al linguaggio.

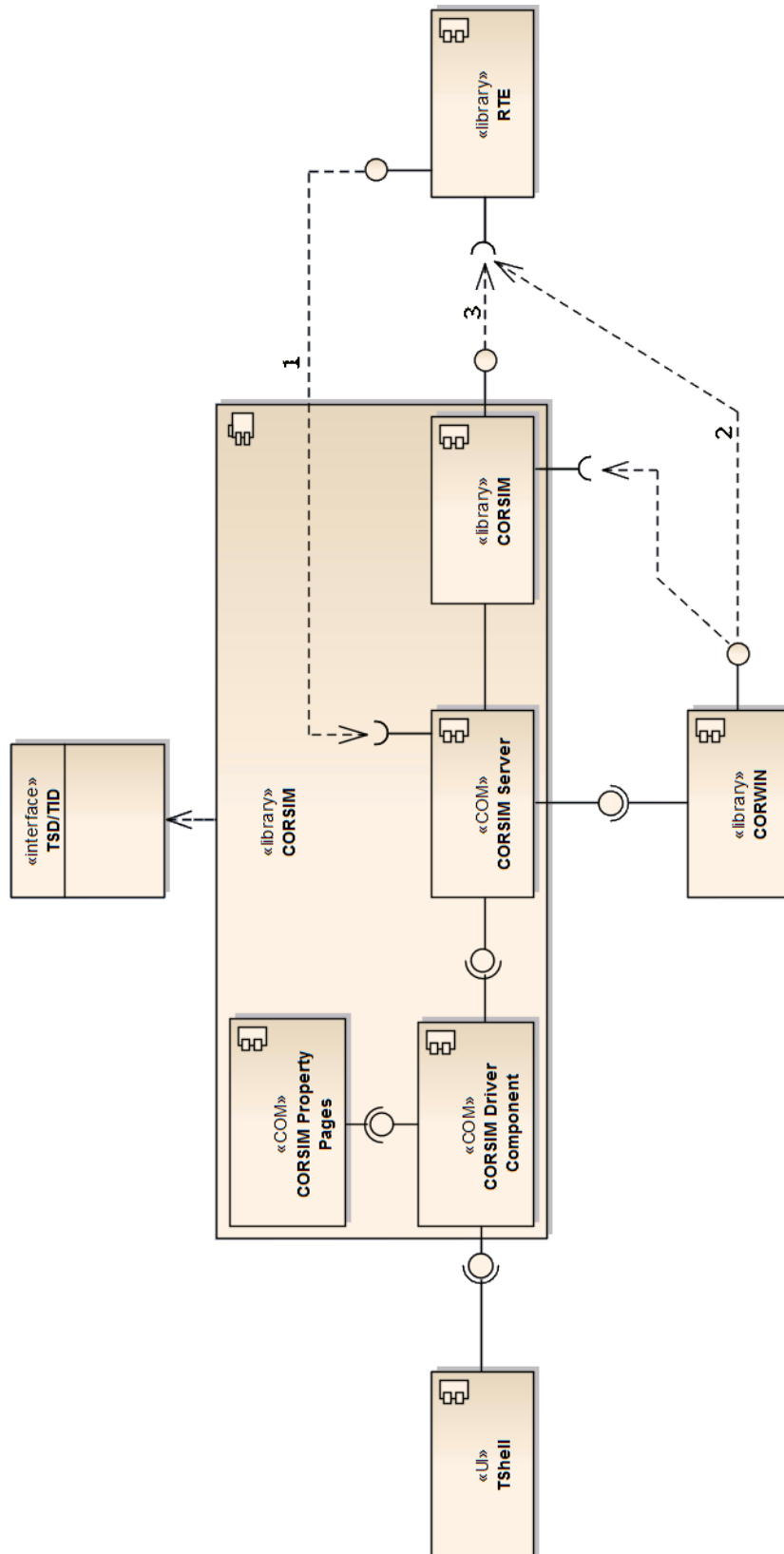


Figura 5.3: Porzione del diagramma dei componenti di TSIS: mostra l'architettura modulare e il funzionamento di CORSIM all'interno di TSIS.

termine CORSIM API al fine di semplificare la discussione. Perciò, la CORSIM API, rappresentata dalla freccia 3 della figura 5.3 nella pagina precedente, oltre a permettere l'estrazione di informazioni relative alla simulazione, permette alla RTE di controllare, eventualmente, molti aspetti della simulazione operata da CORSIM (e. g., aborto della simulazione).

Si osservi che, anche se la figura 5.3 nella pagina precedente non evidenzia tale possibilità, l'architettura di CORSIM supporta l'utilizzo di più RTE contemporaneamente.

Nota 5.2.1. Un attento osservatore noterà come CORSIM, la libreria finalizzata al processo di simulazione, sia a sua volta una RTE automaticamente collegata ai moduli CORSIM Server e CORWIN. Inoltre, la figura 5.4 nella pagina seguente fa notare come tutti i componenti (elencati e descritti nella sottosezione 5.1.1 a pagina 45) dell'ambiente di sviluppo TSIS siano anch'essi dei moduli architetturealmente uguali alle RTE.

5.2.3 Ciclo di vita di CORSIM

Di seguito si presenta il ciclo di vita di CORSIM descrivendo i punti di chiamata che esso esporta tramite apposite funzioni affinché una RTE, implementando ed esportando una funzione per almeno uno di essi, possa interfacciarsi con il processo di simulazione. Tali informazioni sono quindi relative alla API rappresentata dalla freccia 1 nella figura 5.3 nella pagina precedente. Le modalità di implementazione e utilizzo in C++ di tale API sono illustrate nella sottosezione 5.2.5 a pagina 58.

La tabella 5.2 a pagina 57 descrive tutti i punti di chiamata relativi alla linea di esecuzione temporale di CORSIM. Una volta compilata la RTE e ottenuto il relativo file DLL, ognuno dei punti di chiamata di CORSIM deve essere associato alla rispettiva funzione implementata dalla RTE. La sezione che segue descrive in maggior dettaglio tale processo.

5.2.4 Collegare una RTE a CORSIM

Questa sottosezione illustra i passi necessari a espletare il processo di collegamento (i. e., *linking*) di una RTE a CORSIM. Tale processo è eseguibile direttamente tramite TShell.

A scopo esemplificativo si descrive come aggiungere la RTE per il rilevamento e tracciamento del passaggio dei veicoli sui sensori (descritta nella sezione 5.3 a pagina 60). Tale operazione viene svolta tramite il menù *Tools* di TShell scegliendo la voce *Tool Configuration* e cliccando sul pulsante per l'aggiunta (i. e., *Add*) di una RTE. La figura 5.4 nella pagina seguente mostra lo strumento di configurazione degli strumenti appartenenti all'ambiente di TSIS.

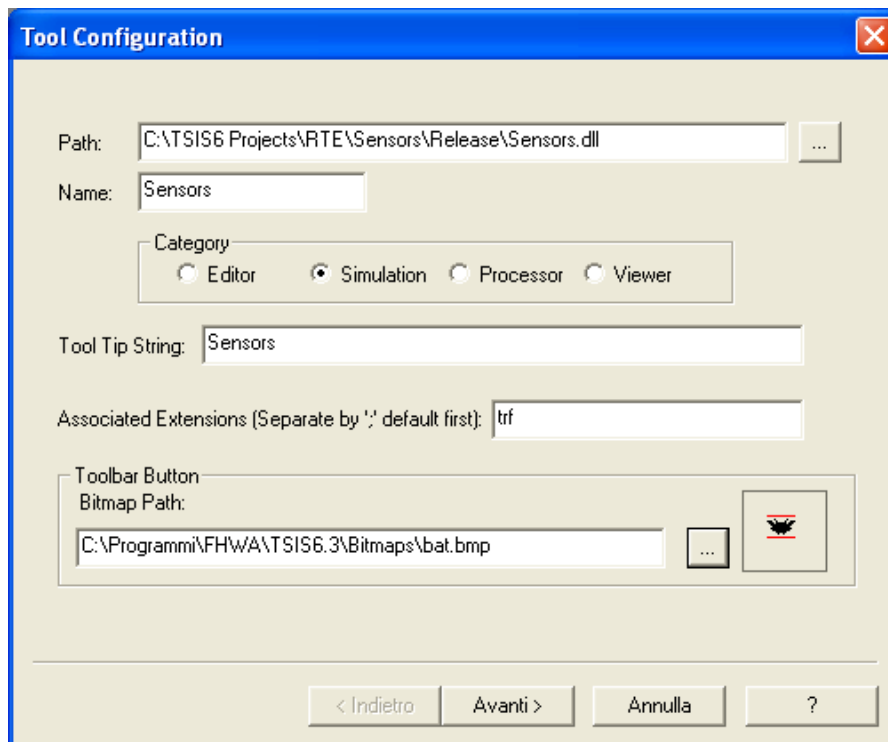


Figura 5.4: Strumento finalizzato all'aggiunta di una RTE a TSIS.

Specificato il percorso a cui risiede il file DLL della RTE, il tipo di RTE, il nome e l'icona che si intende assegnare a tale strumento e il tipo di file a cui va associato (e. g., TRF), la RTE è aggiunta a TSIS. La figura 5.5 mostra la barra degli strumenti di TShell quando un file TRF viene aperto: essa contiene il pulsante per l'avvio della RTE appena aggiunta all'ambiente di sviluppo TSIS.



Figura 5.5: Barra degli strumenti di TShell contenente il pulsante per l'invocazione della RTE aggiunta all'ambiente di sviluppo TSIS.

Tuttavia, come detto, le funzioni della RTE devono essere collegate ai punti di chiamata di CORSIM affinché la RTE risulti completamente funzionante. Per adempiere tale operazione è necessario utilizzare nuovamente lo strumento di configurazione cliccando sul pulsante per la modifica (i. e., *Edit*) di una RTE. A questo punto, selezionando la scheda relativa alle RTE, è possibile invocare lo strumento per effettuare il succitato collegamento. La figura 5.6 nella pagina seguente mostra il processo di collegamento tra le funzioni della RTE e i punti di chiamata di CORSIM.

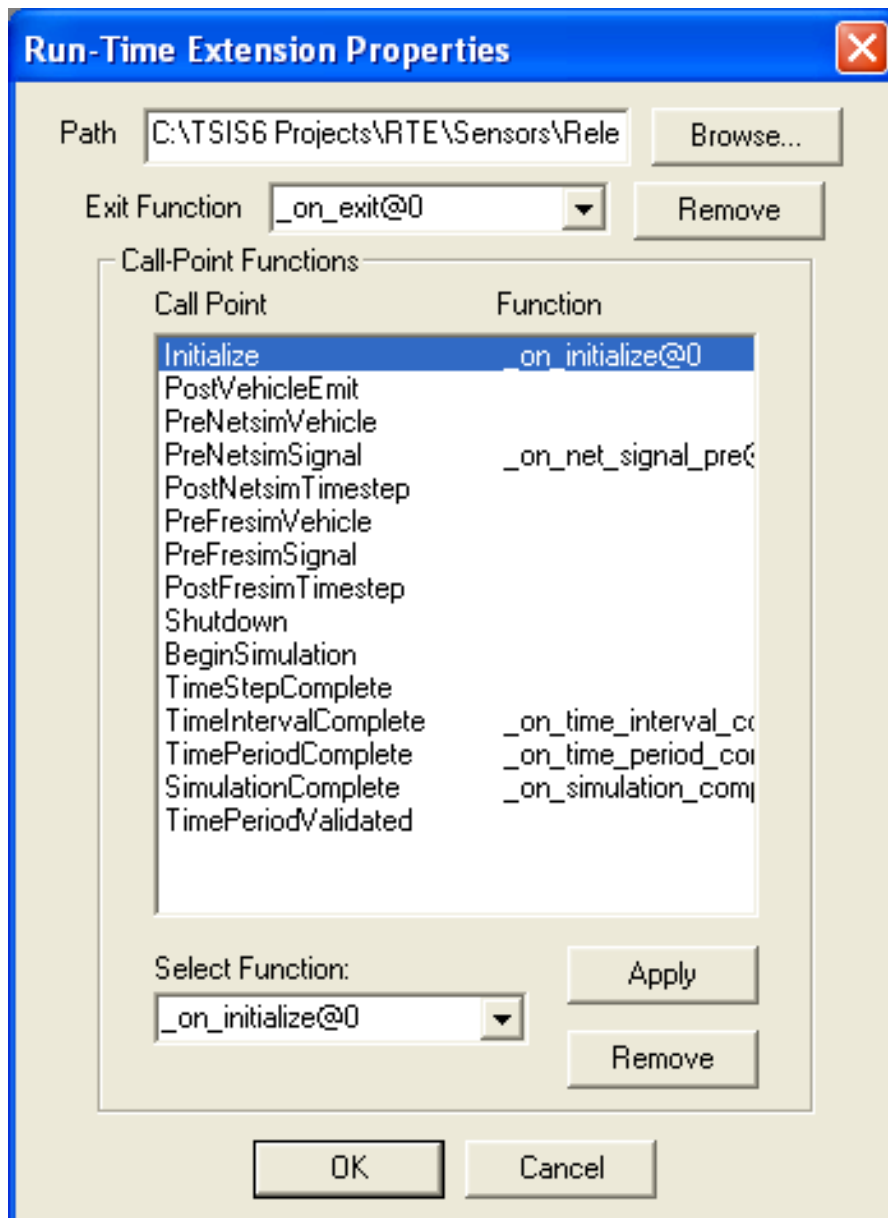
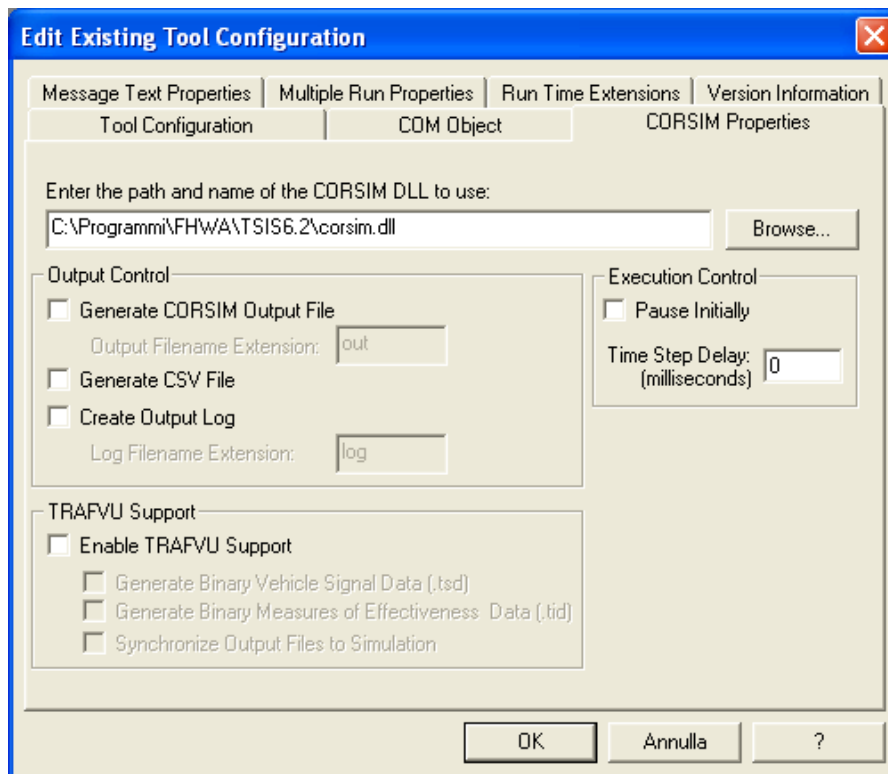
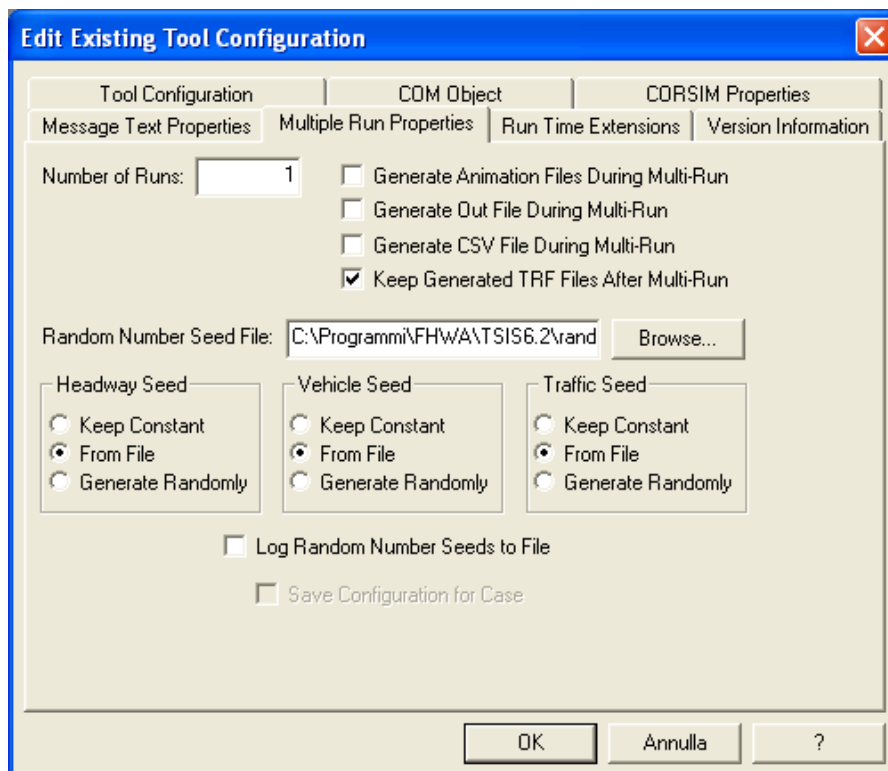


Figura 5.6: Collegamento delle funzioni della RTE ai rispettivi punti di chiamata di CORSIM.

Inoltre, può essere necessario dover configurare la RTE aggiunta in base alle sue esigenze, così come modificare alcune funzionalità di CORSIM relativamente ad essa. Ad esempio, la RTE per il monitoraggio e il tracciamento del passaggio dei veicoli sui sensori non necessita che i file di output di CORSIM vengano generati, né che vengano generati i file per la visualizzazione animata della simulazione in TRAFED. Inoltre, tale RTE, non necessita che la simulazione CORSIM sia eseguita più volte. La figura 5.7 nella pagina successiva mostra la configurazione di tali opzioni.



(a) Proprietà di CORSIM.



(b) Impostazioni proprietà d'esecuzione di CORSIM.

Figura 5.7: Configurazione delle proprietà di CORSIM per la RTE: disattivazione dell'output di CORSIM, della generazione dell'input per TRAFED e delle esecuzioni multiple della simulazione.

Punto di chiamata	Descrizione
Initialize	Chiamato all'inizio della simulazione prima della fase di inizializzazione CORSIM ma dopo la lettura del file TRF di input.
PostVehicleEmit	Chiamato ad ogni <i>time step</i> dopo che i veicoli sono stati immessi nella rete stradale.
PreNetsimVehicle	Chiamato ad ogni <i>time step</i> appena prima che i veicoli inizino a muoversi nel sotto-modello NETSIM.
PreNetsimSignal	Chiamato ad ogni <i>time step</i> appena prima che i segnali (e.g., semafori) del sotto-modello NETSIM vengano impostati.
PostNetsimTimestep	Chiamato in corrispondenza della fine del processo di simulazione di ogni <i>time step</i> di NETSIM e prima che FRESIM inizi a simulare il suo relativo sotto-modello.
PreFresimVehicle	Chiamato ad ogni <i>time step</i> appena prima che i veicoli inizino a muoversi nel sotto-modello FRESIM.
PreFresimSignal	Chiamato ad ogni <i>time step</i> appena prima che i segnali (e.g., semafori) del sotto-modello FRESIM vengano impostati.
PostFresimTimestep	Chiamato in corrispondenza della fine del processo di simulazione di ogni <i>time step</i> di FRESIM.
BeginSimulation	Chiamato dopo l'inizializzazione di CORSIM (i.e., rete stradale piena) in corrispondenza dell'inizio della simulazione.
TimeStepComplete	Chiamato in corrispondenza della fine del processo di simulazione di ogni <i>time step</i> .
TimeIntervalComplete	Chiamato in corrispondenza della fine del processo di simulazione di ogni <i>time interval</i> .
TimePeriodComplete	Chiamato in corrispondenza della fine del processo di simulazione di ogni <i>time period</i> .
TimePeriodValidated	Chiamato in corrispondenza della fine del processo di lettura e validazione del file di input relativo a ogni <i>time period</i> , prima dell'inizializzazione e dell'effettivo inizio del processo di simulazione di ogni <i>time period</i> .
SimulationComplete	Chiamato in corrispondenza della fine della simulazione e prima della completa terminazione del processo.
Shutdown	Chiamato appena prima che l'esecuzione di CORSIM termini.
Exit	Chiamato in corrispondenza della fine dell'intero processo di simulazione.

Tabella 5.2: Descrizione di punti di chiamata che CORSIM espone all'esterno.

5.2.5 Utilizzo delle API

Lungi dal volere fornire una documentazione esaustiva delle API dell'ambiente di sviluppo TSIS, in questa sezione, si intende presentare, tramite esempi, le modalità di utilizzo di tali API in linguaggio C++.

Ad esempio, il listato 5.1 mostra il codice di intestazione necessario a definire ed esportare la funzione `on_initialize` (listato 5.2), che, come mostrato dalla figura 5.6 a pagina 55 viene collegata al punto di chiamata `Initialize` di `CORSIM`. Si osservi che la scelta del nome di tale funzione non è vincolata ad alcun criterio.

```

1 #ifdef __cplusplus
2 #define DLL_EXPORT extern "C" __declspec(dllexport)
3 #endif // __cplusplus

```

Sorgente 5.1: Costrutto per l'esportazione delle funzioni RTE

```

1 DLL_EXPORT void __stdcall on_initialize() {
2     // implementation
3 }

```

Sorgente 5.2: Esempio di funzione RTE esportata

Il listato 5.3 illustra invece come importare una delle funzioni esposte dalla `CORWIN` API in una RTE. Nello specifico, l'esempio in questione, è relativo all'importazione della funzione `OutputString`, finalizzata all'invio di messaggi di output a `TShell` durante la simulazione `CORSIM`.

```

1 #ifdef __cplusplus
2 #define CORWINAPI extern "C" __declspec(dllimport)
3 CORWINAPI void __stdcall OutputString(const char *str,
4     unsigned int size, int msgCode, unsigned long color);
5 #endif // __cplusplus

```

Sorgente 5.3: Importazione delle `CORWIN` API

Infine, come già detto, `CORSIM` permette l'accesso alla maggior parte dei dati che esso manipola durante la simulazione, esportandoli nella memoria condivisa. Tali dati sono dei seguenti tipi:

1. variabili scalari (non array)
2. array allocati staticamente
3. array allocati dinamicamente
4. funzioni esposte tramite API.

Il listato 5.4 mostra il costrutto utilizzabile per l'importazione di dati e funzioni dalla cosiddetta CORSIM API.

```

1 #ifdef __cplusplus
2 #define DLL_IMPORT extern "C" __declspec(dllimport)
3 #endif // __cplusplus

```

Sorgente 5.4: Costrutto per l'importazione delle CORSIM API

Tale costrutto viene poi utilizzato per l'effettiva importazione di dati e funzioni da CORSIM. Il listato 5.5 riporta un esempio di importazione per ogni tipo di dati che le CORSIM API rendono disponibile:

1. a linea 2 si importa il numero di sensori presenti sulla rete stradale urbana, esportato da CORSIM tramite la variabile scalare NETSIM_DETECTORS_mp_NUMDET e rinominato, a linea 3, in net_det_num
2. a linea 5 si importa la lista statica (i.e., di dimensione massima prefissata) dei numeri identificativi assegnati dall'utente ai nodi della rete stradale, SIN075.NMAP; rinominata a linea 9 in net_node_num
3. a linea 11 si importa la lista dinamica delle informazioni sui sensori, *NETSIM_DETECTORS_mp_DETMOD; rinominata, a linea 12, in net_det_mod
4. a linea 14 si importa abortcorsim, una funzione delle CORSIM API finalizzata al controllo dell'esecuzione della simulazione da parte della RTE

```

1 // netsim scalar non-array variables
2 DLL_IMPORT int NETSIM_DETECTORS_mp_NUMDET;
3 #define net_det_num NETSIM_DETECTORS_mp_NUMDET
4 // netsim statically allocated arrays
5 DLL_IMPORT struct
6 {
7     int NMAP[8999];
8 } SIN075;
9 #define net_node_num SIN075.NMAP
10 // netsim dynamically allocated arrays
11 DLL_IMPORT int *NETSIM_DETECTORS_mp_DTMODD;
12 #define net_det_id NETSIM_DETECTORS_mp_DTMOD
13 // netsim exported functions
14 DLL_IMPORT void __stdcall abortcorsim(void);

```

Sorgente 5.5: Importazione di oggetti delle CORSIM API

5.3 ESTENSIONE

Avendo presentato l'ambiente di sviluppo TSIS e le API che esso fornisce per la sua estensione, è ora possibile descrivere l'estensione a tempo d'esecuzione sviluppata al fine di generare i succitati dataset.

Come detto nella sottosezione 5.1.2 a pagina 47, una delle principali limitazioni di CORSIM consiste nell'impossibilità di ottenere dei dati non aggregati dal processo di simulazione. Ciò poiché il minimo intervallo temporale che CORSIM permette di utilizzare è pari a 1 secondo per le reti stradali urbane e al più 0.1 secondi per le reti stradali extraurbane. Perciò, è emerso il problema di sorpassare tale limitazione. La RTE sviluppata, Sensors DLL, risponde a tale necessità monitorando determinati elementi (i. e., i sensori) di un qualsiasi tipo di rete stradale e tracciando gli eventi ad essi correlati (i. e., passaggio di un veicolo) su un file di output esterno a TSIS. Lo scopo di questa sezione consiste nel presentare il funzionamento di Sensors DLL.

5.3.1 Sensors DLL

L'obiettivo ultimo di Sensors DLL consiste nella generazione di un file CSV³⁴ che rappresenti il passaggio dei veicoli sui vari sensori presenti nella rete stradale nel tempo. Il monitoraggio dei sensori deve essere effettuato con la massima granularità temporale possibile (i. e., 0.1 secondi), anche nel caso di reti stradali urbane.

A tale scopo si è utilizzato il dato NETSIM_DETECTORS_mp_DETON, rinominato in `net_det_on`, esportato da CORSIM nella memoria condivisa. Tale campo indirizza un array dinamico la cui lunghezza è pari al numero di sensori sulla rete stradale. Ognuno degli elementi di tale array è costituito da 10 bit: l'*i*-esimo bit rappresenta l'attivazione (i. e., 1) o meno del relativo sensore nell'*i*-esimo passo temporale minimo (i. e., 0.1 secondi). Ogni elemento rappresenta perciò il passaggio dei veicoli sul sensore durante un intervallo temporale fisso di 1 secondo.

Di seguito si presentano le operazioni principali che Sensors DLL effettua:

1. ottenere il nome del file TRF di input, rappresentante la rete stradale e il modello di simulazione
2. effettuare il *parsing*³⁵ di tale file creando gli oggetti relativi a ogni elemento (e. g., intersezioni, strade, sensori) della rete stradale
3. rilevare il passaggio dei veicoli sui sensori ogni secondo

³⁴ Comma Separated Values (CSV) è un formato basato su file di testo utilizzato per l'importazione ed esportazione (ad esempio da fogli elettronici o database) di una tabella di dati.

³⁵ Il *parsing* consiste nel processo atto ad analizzare un input in modo da determinare la sua struttura grammaticale grazie ad una data grammatica formale.

4. ricostruire l'intero flusso di veicoli su ogni sensore durante tutto il tempo di simulazione
5. creare un file di output che contenga le informazioni ottenute.

Le operazioni 1 e 2 vengono compiute in corrispondenza dell'inizializzazione di CORSIM e quindi della RTE. Il risultato di tali operazioni è un insieme di istanze correlate rappresentanti gli elementi della rete stradale di input e le caratteristiche di ognuno di essi. Il diagramma delle classi di Sensors DLL, mostrato in figura 5.8 a pagina 63, illustra le relazioni di associazione e aggregazione degli oggetti con cui si è scelto di rappresentare le reti stradali TSIS. La classe CNetwork rappresenta la rete stradale, composta da un insieme di intersezioni e strade, elementi rappresentati rispettivamente dalle classi CNode e CLink. Ogni strada può a sua volta essere composta da più corsie, elementi rappresentati tramite la classe CLane, e contenere dei sensori, elementi rappresentati tramite la classe CDetector. Inoltre, poiché è possibile che alcuni sensori siano posti esclusivamente su una corsia piuttosto che su tutta la superficie della strada, sussiste una relazione anche fra la classe rappresentante le corsie e quella rappresentante i sensori. Invece, la classe CBinary non rappresenta alcun elemento concreto della rete stradale: la sua funzione è esclusivamente quella di incapsulare l'intero `net_det_on` e convertirlo nella corretta sequenza di bit rappresentante il flusso dei veicoli su un sensore. La procedura che effettua tali operazioni di inizializzazione della rete stradale nella RTE, chiamata `on_initialize`, è collegata al punto di chiamata `Initialize`, così come mostrato dalla figura 5.6 a pagina 55.

Configurata la rete stradale, Sensors DLL può monitorare i sensori di pari passo con l'esecuzione della simulazione da parte di CORSIM. Tale operazione viene effettuata ad ogni intervallo temporale di NETSIM poiché la procedura che la incorpora, chiamata `on_net_signal_pre`, è collegata al punto di chiamata `PreNetsimSignal` di CORSIM. Il listato 5.6 a pagina 62 illustra una versione semplificata del metodo C++ preposto all'esecuzione di tale procedura. Essa consiste nell'iterazione della lista di sensori afferenti ad una strada: per ogni sensore (ciclo a linea 8), recuperato l'identificatore che CORSIM utilizza per rappresentarlo (istruzione a linea 10), si ottiene il relativo elemento dell'array `net_det_on` (istruzione a linea 11), un intero rappresentante il flusso di veicoli sul sensore nell'ultimo secondo di simulazione. Tale intero viene poi convertito nella corretta sequenza di 10 bit tramite la classe CBinary a linea 12. Da linea 13 a linea 32 si itera in ordine inverso la sequenza di bit al fine di estrapolare e memorizzare lo stato (i.e., 1 in caso di veicolo rilevato, 0 altrimenti) del sensore in ogni passo temporale minimo (i.e., 0.1 secondi). Quindi questa procedura, ripetuta per tutte le strade presenti nella rete stradale e ad ogni intervallo temporale, memorizza per ogni sensore una lista di valori booleani.

Completata la simulazione e di conseguenza anche la procedura di monitoraggio dei sensori, Sensors DLL, in corrispondenza del punto di chiamata SimulationComplete di CORSIM, genera un file CSV in cui memorizza il tempo, il *time period* della rilevazione e la dinamica di stato di ogni sensore. La sottosezione 5.3.2 nella pagina seguente si occupa di presentare in maggior dettaglio l'output di Sensors DLL.

```

1 void CLink::processDetectors(void) {
2     int det, new_state, old_state = 0;
3     POSITION pos, pos_i = NULL;
4     CDetector *detector = NULL;
5     CInteger *pi = NULL;
6     CBinarySequence *sequence = NULL;
7     pos = m_detector_list.GetHeadPosition();
8     while (pos != NULL) {
9         detector = m_detector_list.GetNext(pos);
10        int det_num = detector->getCorsimId();
11        det = net_det_on[det_num];
12        sequence = CBinarySequence::convert(det);
13        pos_i = sequence->sequence.GetTailPosition();
14        old_state = detector->getState() ? 1 : 0;
15        for (index = 0; index < 10; index++) {
16            pi = sequence->sequence.GetPrev(pos_i);
17            new_state = pi->data;
18            if ((old_state == 0) && (new_state == 1)) {
19                detector->setState(true, !is_init);
20            }
21            if ((old_state == 1) && (new_state == 1)) {
22                float atime = detector->getActivationTime();
23                detector->setState(true, !is_init);
24            }
25            if ((old_state == 1) && (new_state == 0)) {
26                detector->setState(false, !is_init);
27            }
28            if ((old_state == 0) && (new_state == 0)) {
29                detector->setState(false, !is_init);
30            }
31            old_state = new_state;
32        }
33        delete sequence;
34    }
35 }

```

Sorgente 5.6: Metodo della classe CLink per la rilevazione del passaggio dei veicoli sui sensori

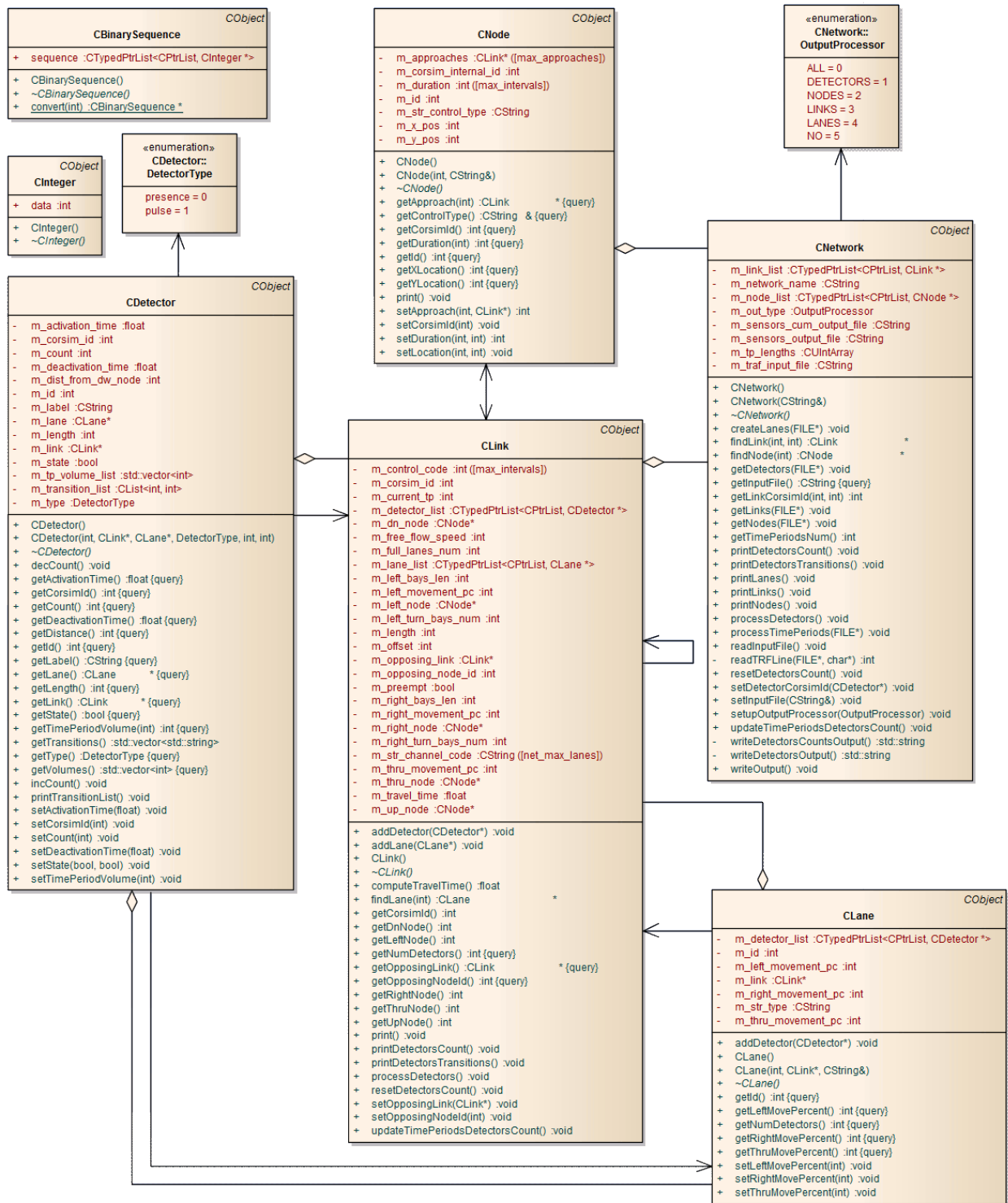


Figura 5.8: Diagramma delle classi di Sensors DLL.

5.3.2 Formato dell'output

Di seguito si presenta un esempio di file CSV di output generato da Sensors DLL. Come anticipato, esso codifica l'onda quadra di ogni

seniore a spira magnetica posto sulla rete stradale.

```
time,tp,D131,D231,D232,D211,D212,D213
0.10000,1,0,0,0,0,0,0
0.20000,1,0,0,1,0,0,0
0.30000,1,0,0,1,0,0,0
0.40000,1,0,0,0,0,0,0
0.50000,1,0,0,0,0,1,0
0.60000,1,0,0,0,0,1,0
0.70000,1,0,0,0,0,0,0
0.80000,1,0,0,0,0,0,0
0.90000,1,0,0,0,0,0,0
1.00000,1,0,1,0,0,0,0
1.10000,2,0,1,0,0,0,1
1.20000,2,0,1,0,0,0,1
1.30000,2,0,1,0,0,0,0
1.40000,2,0,0,0,0,0,0
1.50000,2,0,0,1,0,0,0
1.60000,2,0,0,1,0,0,0
1.70000,2,0,0,0,0,0,0
1.80000,2,1,0,0,0,0,0
1.90000,2,1,0,0,0,0,0
2.00000,2,1,0,0,0,0,0
```

Sorgente 5.7: Formato di output di Sensors DLL

La tabella 5.3 chiarisce il significato di ogni colonna dei dati tabulari restituiti da Sensors DLL.

Nome colonna	Descrizione
time	Tempo di simulazione a cui è stato effettuato il monitoraggio dei sensori.
tp	Indice del corrente periodo temporale (i. e., <i>time period</i>).
identificatore sensore	1 in caso di presenza di un veicolo sul rispettivo sensore, 0 altrimenti.

Tabella 5.3: Descrizione della semantica dei file CSV generati da Sensors DLL.

Si osservi che, il fatto che il sensore D232, in corrispondenza dell'istante 1.6 (5^a colonna) abbia valore 1 non indica che il veicolo sia stato rilevato esattamente in tale istante, bensì ciò indica che durante l'intervallo temporale [1.5, 1.6] tale sensore è stato attivato dal passaggio di un veicolo.

5.4 APPLICATIVI DI SUPPORTO

Al fine di automatizzare e facilitare la creazione di dataset relativi al traffico tramite la RTE trattata nel corrente capitolo, si è sviluppato un insieme di strumenti dediti alla manipolazione dei file di output di Sensors DLL.

Di seguito si elencano le operazioni di manipolazione che tali strumenti supportano:

1. sostituzione (e eventualmente rimozione) della colonna relativa ai periodi temporali con una nuova colonna che rappresenti la classe di un insieme di osservazioni; tale colonna è automaticamente generata in base a un sistema di regole (e.g., *matching* tra periodo temporale e classe).
2. partizionamento del file di output di Sensors DLL in più file in base a vincoli temporali (e.g., divisione del file in blocchi di 60 secondi ciascuno)
3. ottimizzazione del file tramite rimozione delle linee duplicate (i.e., linee in cui non è avvenuto alcun cambiamento di stato dei sensori).

Lo scopo di questo capitolo è presentare i risultati ottenuti dalla sperimentazione degli algoritmi di classificazione e apprendimento strutturale oggetto di questo lavoro di tesi (si vedano rispettivamente i capitoli 2 e 3 a pagina 21 e a pagina 33).

La generazione di dataset inerenti il traffico automobilistico è stata effettuata, come detto, tramite un software commerciale apposito, TSIS, e una estensione a tempo d'esecuzione, Sensors DLL, sviluppata a tale fine. Per maggiori dettagli riguardanti questi strumenti si rimanda al capitolo 5 a pagina 42.

Questo capitolo è quindi articolato in due parti, una per ogni dataset su cui è stata effettuata la succitata sperimentazione. Ognuna di tali sezioni è a sua volta suddivisa in due sottosezioni: la prima finalizzata alla descrizione del modello TSIS da cui è stato generato il dataset oggetto della sperimentazione e alla descrizione di tale dataset, e la seconda finalizzata alla presentazione dei risultati di tale sperimentazione.

Al fine di valutare l'accuratezza dei metodi predittivi sviluppati, la sperimentazione è stata eseguita utilizzando la tecnica chiamata *cross-validation*. Tale metodologia di validazione è infatti pensata per accertare la capacità di generalizzazione di un modello statistico di tipo predittivo su un insieme di dati non conosciuto. Nello specifico, entrambe le sperimentazioni, e entrambi i metodi predittivi valutati (i.e., classificazione e apprendimento strutturale), sono state condotte tramite una *k-fold cross-validation* con $k = 10$.

6.1 DATASET #1

Lo scopo di questa sezione è presentare il dataset #1 e i risultati di classificazione e apprendimento strutturale ottenuti utilizzando tale insieme di *dati completi* come input.

6.1.1 Modello TSIS

Al fine di introdurre il dataset #1 è necessario presentare la rete stradale TSIS e il relativo modello di simulazione CORSIM da cui esso è stato generato.

La figura 6.1 nella pagina successiva mostra la rete stradale che compone il modello TSIS oggetto di questa sottosezione.

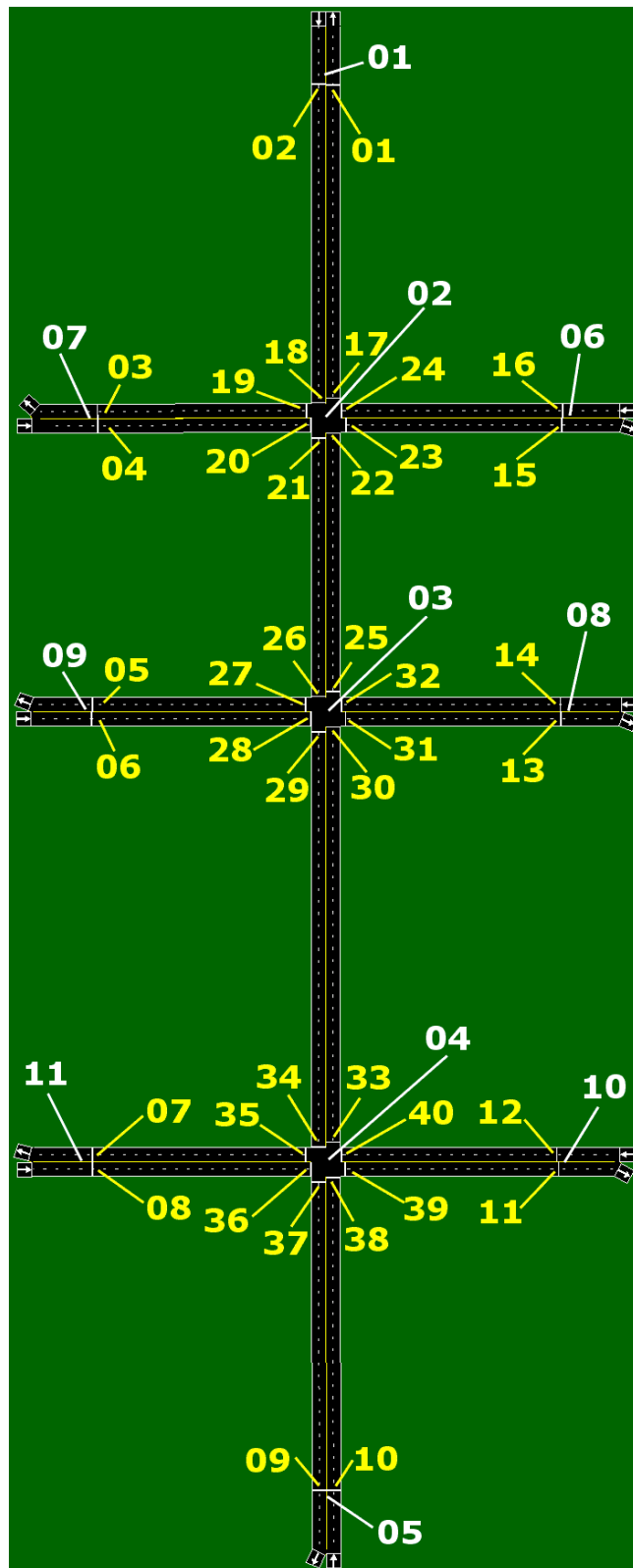


Figura 6.1: Visualizzazione del file TNO rappresentante la rete stradale da cui viene generato il dataset #1. I sensori sono etichettati tramite degli indici numerici di colore giallo mentre i nodi sono etichettati tramite degli indici numerici di colore bianco.

Tale figura evidenzia:

- la disposizione dei sensori (indicati tramite degli indici numerici di colore giallo) utilizzati per il monitoraggio del passaggio dei veicoli
- le strade, tutte composte da 2 corsie, e i nodi (indicati tramite degli indici numerici di colore bianco) da cui è composta la rete stradale
- i sensi di marcia delle corsie

Di seguito si riportano le ulteriori caratteristiche di tale rete stradale:

- la larghezza delle corsie è 3,65 m
- la lunghezza della strada principale (i.e., da nodo 01 a nodo 05) è 780 m mentre la lunghezza delle strade ad essa perpendicolari è pari a 300 m
- la dimensione dei sensori è 0,30 m, da cui consegue che la dimensione della zona sensibile (in tutte le direzioni) è 2,13 m
- la distanza dei sensori dal nodo più vicino è 6,5 m

Le intersezioni principali della rete stradale (i.e., nodi con indici 02, 03 e 04 nella figura 6.1 nella pagina precedente) sono controllate da semafori il cui ciclo di controllo dura 100 secondi. La figura 6.2 illustra la configurazione del piano semaforico per ogni intersezione controllata durante tutto il tempo di simulazione.

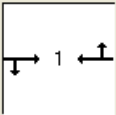
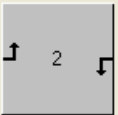
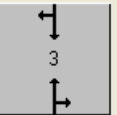

Phase Movements:				
Green Time:	30	20	20	10
Yellow Time:	1	1	1	1
All Red Time:	4	4	4	4

Figura 6.2: Piano semaforico predefinito delle intersezioni 02, 03 e 04 della rete stradale TSIS relativa al dataset #1.

Come anticipato, ogni strada che compone la rete stradale contiene, nei pressi dei nodi (sia che essi siano delle intersezioni, sia che essi siano dei nodi di ingresso o di transito), dei sensori finalizzati al rilevamento e tracciamento del passaggio dei veicoli: essi sono indicati nella figura 6.1 nella pagina precedente da indici numerici di colore giallo. La tabella 6.1 nella pagina successiva elenca il nome reale di tali sensori, utilizzato per la generazione del relativo dataset (a tal riguardo, un esempio di dataset è stata riportato nella sottosezione 5.3.2 a pagina 63).

#	Sensore	#	Sensore	#	Sensore	#	Sensore
01	D212	11	D4102	21	D231	31	D381
02	D121	12	D1041	22	D322	32	D832
03	D272	13	D382	23	D261	33	D431
04	D721	14	D831	24	D622	34	D342
05	D392	15	D262	25	D321	35	D4111
06	D931	16	D621	26	D232	36	D1142
07	D4412	17	D211	27	D391	37	D451
08	D1141	18	D122	28	D932	38	D542
09	D452	19	D271	29	D341	39	D4101
10	D541	20	D722	30	D432	40	D1042

Tabella 6.1: Corrispondenza fra gli identificatori dei sensori del dataset #1 e l'indice con cui essi sono indicati nella figura 6.1.

Il modello di simulazione CORSIM creato per questa rete stradale prevede un tempo di simulazione totale pari a 24 ore, suddiviso in 16 *time period*, ognuno dei quali composto da *time interval* della durata di 1 secondo: la tabella 6.2 illustra la durata (in ore) di ognuno di essi e introduce la corrispondenza fra essi e la fase della giornata, rappresentata come la variabile classe nel relativo dataset.

Fascia oraria	Time Period	Fase	Classe
07:00 - 08:00	1	mattino	1
08:00 - 09:00	2	mattino	1
09:00 - 10:00	3	mattino	1
10:00 - 12:00	4	giorno	2
12:00 - 14:00	5	giorno	2
14:00 - 16:00	6	giorno	2
16:00 - 17:00	7	giorno	2
17:00 - 18:00	8	pomeriggio	3
18:00 - 19:00	9	pomeriggio	3
19:00 - 20:00	10	pomeriggio	3
20:00 - 22:00	11	sera	4
22:00 - 24:00	12	sera	4
00:00 - 02:00	13	notte	5
02:00 - 04:00	14	notte	5
04:00 - 05:00	11	notte	5
05:00 - 07:00	16	alba	6

Tabella 6.2: Caratterizzazione dei periodi temporali (i.e., *time period*) del modello TSIS relativo al dataset #1.

Ogni fase della giornata (i.e., classe) è caratterizzata da un andamento diverso del traffico:

- al mattino traffico intenso su tutta la rete stradale e code in

direzione del nodo 05

- durante il giorno traffico moderato su tutta la rete e code residue verso il nodo 05 nella prima parte di tale fase
- nel pomeriggio traffico intenso su tutta la rete stradale e code in direzione del nodo 01
- durante la sera traffico moderato su tutta la rete stradale e code residue verso il nodo 01 nella prima parte di tale fase
- durante la notte pochi veicoli in tutte le direzioni
- all'alba aumento dei veicoli diretti verso il nodo 05

L'andamento del traffico è modellato in CORSIM variando i flussi di ingresso dei veicoli e le percentuali di svolta nelle varie intersezioni al variare dei periodi temporali o, in generale, delle fasi di simulazione previste. La tabella 6.3 illustra i flussi di ingresso nella rete stradale, espressi come numero di veicoli orari (i.e., vph), per ogni fase della simulazione. Si osservi che l'immissione di veicoli sulla rete stradale avviene seguendo una distribuzione di probabilità di Erlang con parametro *shape* pari a 2.

Fase	01 → 05	05 → 01	Strade secondarie
	vph	vph	vph
mattino	600	200	100
giorno	300	300	50
pomeriggio	200	600	100
sera	100	200	50
notte	50	50	25
alba	200	100	50

Tabella 6.3: Numero di veicoli orari immessi nella rete stradale del dataset dataset #1 al variare della fase di simulazione.

La tabella 6.4 nella pagina successiva illustra le percentuali di svolta al variare della fase di simulazione. L'intestazione di tale tabella indica i nodi di provenienza dei veicoli, il senso di marcia e le direzioni previste dall'intersezione che segue il nodo di partenza. Le celle invece indicano la distribuzione delle svolte effettuate dai veicoli nelle intersezioni.

Si consideri, ad esempio, la prima cella della tabella 6.4: l'80% dei veicoli provenienti dai nodi 01, 02 e 03, giunti alle intersezioni che seguono tali nodi (i.e., nodi 02, 03 e 04, rispettivamente; si veda a tal riguardo la figura 6.1 a pagina 67), proseguono senza svoltare; il 10% dei veicoli provenienti dai nodi 01, 02 e 03, invece, svolterà verso destra; idem verso sinistra.

Nodi	01 - 02 - 03			03 - 04 - 05			06 - 08 - 10			07 - 09 - 11		
	→			←			→			←		
	↔	↔	↓	↔	↔	↓	↘	↔	↓	↖	↔	↓
	%			%			%			%		
mattino	10	10	80	20	20	60	80	10	10	10	80	10
giorno	20	20	60	20	20	60	60	20	20	20	60	20
pomeriggio	20	20	60	10	10	80	10	80	10	80	10	10
sera	40	40	20	10	10	80	30	40	30	40	30	30
notte	33	33	33	33	33	33	33	33	33	33	33	33
alba	10	10	80	40	40	20	40	30	30	30	40	30

Tabella 6.4: Percentuali di svolta dei veicoli per ogni intersezione della rete stradale del dataset #1, al variare della fase di simulazione.

Come detto nella sottosezione 5.3.1 a pagina 60, eseguendo il modello di simulazione CORSIM descritto in questa sezione con Sensors DLL si ottiene un file di output (si veda il listato d'esempio 5.7 a pagina 64) che, adeguatamente processato, compone un dataset sottoponibile agli algoritmi di apprendimento e classificazione presentati nei capitoli 2 e 3.

Di seguito si descrivono le caratteristiche di tale dataset:

- tempo totale pari a 86400 secondi (i. e., 24 ore)
- suddivisione in 864 elementi (i. e., file)
- ogni elemento include 100 secondi di traiettorie di ogni sensore della succitata rete stradale
- ogni elemento è stato ottimizzato, cioè sottoposto all'eliminazione delle righe in cui non si verifica alcuna transizione di stato dei sensori
- la colonna `tp` è stata rimossa e sostituita con la colonna `class` in base alla corrispondenza riportata nella tabella 6.2 a pagina 69
- l'intestazione di ogni elemento è:
`time, class, D122, D121, D341, D342, D722, D721, D1142, D1141, D231, D232, D541, D542, D4112, D4111, D431, D432, D4101, D4102, D451, D452, D832, D831, D261, D262, D321, D322, D392, D391, D211, D212, D272, D271, D1042, D1041, D622, D621, D932, D931, D381, D382.`

6.1.2 Risultati

6.2 DATASET #2

Lo scopo di questa sezione è presentare il dataset #1 e i risultati di classificazione e apprendimento strutturale ottenuti utilizzando tale insieme di *dati completi* come input. A tale scopo, nella sottosezione

che segue, si descrive la rete stradale TSIS e il relativo modello di simulazione CORSIM da cui tale dataset è stato generato.

6.2.1 Modello TSIS

Il modello TSIS in questione, a differenza di quello presentato nella sottosezione 6.1.1, riproduce una rete stradale reale. Nello specifico esso riproduce la rete stradale circostante *Viale Cesare Battisti* (20900 *Monza, MB - Italia*). Disponendo di una serie di informazioni reali, tra cui i dati dei sensori posti su tale rete stradale, si sono potute inferire le caratteristiche della circolazione dei veicoli (e.g., flusso di veicoli in ingresso) e quindi riprodurre fedelmente l'andamento del traffico durante tutti i giorni della settimana tramite degli adeguati modelli di simulazione CORSIM. Nello specifico si sono realizzati 3 modelli CORSIM: essi rappresentano la stessa rete stradale ma diversi modelli di traffico, uno per i giorni lavorativi (dal lunedì al venerdì), uno per il sabato e uno per la domenica.

La figura 6.3 nella pagina seguente illustra la rete stradale oggetto di studio, evidenziando e identificando i nodi e i collegamenti stradali utilizzati per la sua riproduzione in TSIS.

La tabella 6.5 illustra la correlazione tra gli indici dei nodi CORSIM e ciò che essi rappresentano nel modello in questione.

Nodo/i	Descrizione
01	Intersezione <i>Battisti - Rossini - Alighieri</i>
02	Intersezione <i>Battisti - Brianza - Margherita</i>
03	Nodo di ingresso da <i>Viale Brianza</i>
04 - 05 - 06	Intersezione <i>Boccaccio - Margherita</i>
08	Intersezione <i>Battisti - Boito - Tognetti</i>
09	Intersezione <i>Battisti - Donizzetti - Volta</i>
10	Nodo di ingresso da <i>Via Alighieri</i>
11	Nodo di ingresso da <i>Via Rossini</i>
12	Nodo di ingresso da <i>Via Donizzetti</i>
13	Nodo di uscita da <i>Via Volta</i>
14	Nodo di ingresso da <i>Via Tognetti</i>
15	Nodo di ingresso da <i>Via Boito</i>
07 - 16	Diramazione <i>Boito - Battisti</i>
17	Nodo di ingresso da <i>Viale Battisti</i>
18	Nodo di ingresso da <i>Via Boccaccio</i>
19	Nodo di ingresso da <i>Viale Margherita</i>

Tabella 6.5: Caratterizzazione degli identificatori delle intersezioni (o nodi) del dataset #2.

Tutte le intersezioni della rete stradale (i.e., nodi con indici 01, 02, 04, 05, 06, 07, 08, 09 nella tabella 6.5) sono controllate da semafori la cui configurazione riproduce i piani semaforici reali.





Figura 6.4: Visualizzazione del file TNO rappresentante le intersezioni della rete stradale da cui viene generato il dataset #2.

La rete stradale in questione, oltre a riprodurre i sensori realmente esistenti, contiene nei pressi di ognuna delle succitate intersezioni degli ulteriori sensori per la rilevazione del passaggio dei veicoli. Tutti i sensori sono posti su una sola corsia cosicché il flusso di dati da essi generato sia associabile alla canalizzazione della relativa corsia. Di seguito si elencano le informazioni relative ai sensori reali del modello in questione:

5094

Posto nei pressi dell'intersezione 06 (i.e., *Boccaccio - Margherita*), rileva i veicoli provenienti dal nodo 19 (i.e., corsia sinistra di *Viale Margherita*) diretti verso l'intersezione 02 (i.e., *Battisti - Brianza - Margherita*).

5095

Posto nei pressi dell'intersezione 06 (i.e., *Boccaccio - Margherita*), rileva i veicoli provenienti dal nodo 18 (i.e., corsia sinistra *Via Boccaccio*) che svoltano in diagonale a sinistra su *Viale Margherita*.

5096

Posto nei pressi dell'intersezione 02 (i.e., *Battisti - Brianza - Margherita*), rileva i veicoli provenienti dal nodo 03 (i.e., corsia sinistra di *Viale Brianza*) diretti verso il nodo 04 (i.e., *Boccaccio - Margherita*).

5097

Posto nei pressi dell'intersezione 02 (i.e., *Battisti - Brianza - Margherita*), rileva i veicoli su *Viale Margherita* (corsia sinistra) che svoltano a sinistra su *Viale Battisti*.

5098

Posto nei pressi dell'intersezione 02 (i.e., *Battisti - Brianza - Margherita*), rileva i veicoli su *Viale Battisti* (corsia sinistra) che svoltano in diagonale a sinistra su *Viale Brianza*.

5099

Posto nei pressi dell'intersezione 02 (i.e., *Battisti - Brianza - Margherita*), rileva i veicoli su *Viale Battisti* (corsia destra) che svoltano a destra in *Viale Margherita*.

I sensori aggiunti alla rete stradale sono identificati secondo uno schema ben preciso. Il loro identificatore è composto in base al senso di marcia. Nello specifico tale numero è una concatenazione dei seguenti valori: identificatore del nodo di partenza, identificatore del nodo d'arrivo, numero indicante la corsia su cui il sensore è posto (i.e., 0 se è la strada possiede una sola corsia, 1 o 2 se il sensore è posto rispettivamente a destra o a sinistra, 7 o 8 nel caso in cui esso sia posto su ulteriori corsie di destra o sinistra). Ad esempio:

- il sensore identificato da D912 è posto sulla corsia di sinistra della strada che collega l'intersezione "Volta" (i.e., nodo 09) all'intersezione "Dante" (i.e., nodo 01)
- il sensore identificato da D1480 è posto sull'unica corsia di "Via Tognetti" (i.e., nodo 14) nei pressi dell'incrocio con l'intersezione "Boito" (i.e., nodo 08).

#	Sensore	#	Sensore	#	Sensore	#	Sensore
01	D782	09	D892	17	D1091	25	D5096*
02	D781	10	D891	18	D212	26	D321
03	D1687	11	D192	19	D211	27	D241
04	D1681	12	D191	20	D110	28	D242
05	D1682	13	D1290	21	D5098*	29	D5094*
06	D981	14	D912	22	D5099*	30	D642
07	D982	15	D911	23	D5097*	31	D5095*
08	D1480	16	D1092	24	D421	32	D541

Tabella 6.6: Corrispondenza fra gli identificatori dei sensori del dataset #2 e l'indice con cui essi sono indicati nella ...

Nella tabella 6.6, i sensori il cui identificatore è contrassegnato da un asterisco corrispondono ai sensori reali, evidenziati anche nella figura 6.5 a pagina 81 tramite un indice numerico di colore rosso, mentre gli altri sensori sono quelli aggiunti per estendere il modello in questione.

Di seguito si riportano le ulteriori caratteristiche di tale rete stradale:

- la larghezza delle corsie è 3,65 m
- le strade del modello TSIS riproducono esattamente, sia in lunghezza che in forma, la porzione di strada reale che modellano:
 - *Viale Battisti* è lungo 1,33 km
 - *Via Boito* è lunga 308 m
 - *Via Tognetti* è lunga 351 m
 - *Via Donizzetti* è lunga 176 m
 - *Via Volta* è lunga 395 m
 - *Via Rossini* è lunga 141 m
 - *Via Dante* è lunga 422 m
 - *Viale Brianza* è lungo 106 m
 - *Viale Margherita* è lungo 505 m
 - *Via Boccaccio* è lunga 336 m

- la dimensione dei sensori è 0,03 m, da cui consegue che la dimensione della zona sensibile (in tutte le direzioni) è 1,85 m
- la distanza dei sensori dall'intersezione più vicina è 6,5 m
- la velocità media dei veicoli: 67 km/h su *Viale Battisti*, 56 km/h sulle altre strade.

Come anticipato, per questo caso di studio si sono creati 3 modelli di simulazione differenti (corrispondenti quindi a 3 differenti sorgenti TRF), al fine di rappresentare i diversi flussi di traffico durante la settimana.

Il file TRF creato per modellare il traffico sulla rete stradale nei giorni lavorativi prevede un tempo di simulazione totale pari a 24 ore, suddiviso in 12 *time period*, ognuno dei quali composto da *time interval* della durata di 100 secondi: la tabella 6.7 illustra la durata (in ore) di ognuno di essi e introduce la corrispondenza fra essi e la fase della giornata, rappresentata come la variabile classe nel relativo dataset.

Nei modelli di traffico relativi al sabato e alla domenica la corrispondenza tra i *time period*, che rimangono identici a quelli mostrati nella tabella 6.7, e la variabile classe è differente. Ad esempio, il *time period* 1 del modello di traffico relativo al sabato è stato associato alla classe 11 mentre il *time period* 1 del modello di traffico relativo alla domenica è stato associato alla classe 21. Ne deriva che gli identificatori della variabile classe per questi due modelli di traffico sono incrementati di una e due decine rispetto agli identificatori usati per il modello di traffico relativo ai giorni lavorativi.

Fascia oraria	Time Period	Fase	Classe
00:00 - 02:05	1	notte	1
02:05 - 04:10	2	notte	1
04:10 - 06:15	3	alba	2
06:15 - 08:20	4	mattino	3
08:20 - 10:25	5	mattino	3
10:25 - 12:30	6	giorno	4
12:30 - 14:35	7	giorno	4
14:35 - 16:40	8	giorno	4
16:40 - 18:45	9	pomeriggio	5
18:45 - 20:50	10	pomeriggio	5
20:50 - 22:55	11	sera	6
22:55 - 24:00	12	sera	6

Tabella 6.7: Caratterizzazione dei periodi temporali (i.e., *time period*) del modello TSIS relativo al dataset #2 (giorni lavorativi).

Ogni fase della giornata (i.e., classe) è caratterizzata da un andamento diverso del traffico. Di seguito si presenta l'andamento generale del traffico durante i giorni lavorativi:

- al mattino (i.e., periodi temporali 4 e 5) traffico diretto verso il centro di *Monza*, prevalentemente da *Viale Battisti* (i.e., nodo di ingresso 17) e, in misura minore, da *Viale Monza* (i.e., nodo di ingresso 03)
- durante il giorno (i.e., periodi temporali 6, 7 e 8) incremento del traffico verso il centro di *Monza*, soprattutto da *Viale Battisti*, e incremento del traffico fra *Viale Boccaccio* e *Viale Margherita* (i.e., nodi 18 e 19)
- durante il pomeriggio (i.e., periodi temporali 9 e 10) il flusso dei veicoli inizia a invertire la sua direzione verso *Viale Battisti*; aumento del numero di veicoli diretti verso le strade laterali (e.g., *Via Donizzetti*, *Via Alighieri*, *Via Rossini*)
- alla sera (i.e., periodi temporali 11 e 12) traffico proveniente dal centro di *Monza* diretto verso *Viale Battisti* e, al contempo, diminuzione dei veicoli provenienti da *Viale Brianza*
- durante la notte (i.e., periodi temporali 1 e 2) riduzione del traffico su tutta la rete
- all'alba (i.e., periodo temporale 3) pochi veicoli circolanti sulla rete stradale, prevalentemente su *Viale Battisti* dove si verifica un leggero incremento dei veicoli in ingresso.

Come detto, l'andamento del traffico è modellato variando i flussi di ingresso. La tabella 6.8 nella pagina seguente illustra il numero di veicoli orari (i.e., vph) in ingresso sulla rete stradale in relazione a ogni periodo temporale del modello di simulazione. Anche per i modelli di simulazione CORSIM relativi a questa rete stradale si è scelto di utilizzare una distribuzione di probabilità di Erlang per governare l'immissione dei veicoli sulla rete stradale.

Si osservi, inoltre, che al fine di definire il succitato andamento del traffico è stata attuata un'adeguata modellazione delle percentuali di svolta in ogni intersezione. Tuttavia, a causa della complessità del modello, non è possibile riportare le variazioni delle percentuali di svolta di ogni intersezione per ogni periodo temporale.

Per chiarezza si sottolinea infine che i profili di traffico del sabato e della domenica sono stati descritti con flussi di ingresso e percentuali di svolta di diversa entità rispetto a quelli appena descritti.

Eseguendo con Sensors DLL uno qualsiasi dei modelli di simulazione (i.e., giorni lavorativi, sabato o domenica) relativi alla rete stradale presentata, si ottiene il relativo file di output che, processato tramite gli strumenti di supporto precedentemente introdotti (si veda la sezione 5.4 a pagina 65), compone un dataset utile alla classificazione dei profili di traffico.

Time Period	19 vph	18 vph	17 vph	15 vph	14 vph	12 vph	11,10 vph	03 vph
1	150	275	375	5	5	5	5	205
2	25	80	105	5	5	5	5	70
3	96	70	325	5	5	5	5	113
4	350	450	1950	20	50	20	20	1080
5	500	600	1855	20	20	20	20	1325
6	750	500	2100	20	20	20	20	1022
7	600	400	1950	20	20	20	20	1000
8	700	600	2390	20	20	20	20	1050
9	700	510	2410	20	20	20	20	1010
10	700	250	2275	20	12	20	20	980
11	370	190	1565	10	10	10	10	575
12	400	75	975	8	8	10	7	400

Tabella 6.8: Numero di veicoli orari immessi nella rete stradale del dataset #2 (giorni lavorativi) al variare del periodo temporale.

Affinché i dataset in questione siano perfettamente riproducibili si sono utilizzati dei file RNS³⁶, semplici file di testo, opportunamente formattati, da cui CORSIM legge il numero di esecuzioni da effettuare e i semi numerici³⁷ relativi ad alcuni aspetti stocastici della simulazione (e. g., incidenti, tipologia di guida dei conducenti).

Di seguito si riporta il file RNS relativo al modello di traffico per i giorni lavorativi della corrente rete stradale.

5		
32317	22061	28769
19372	19528	8812
10695	338	20744
16724	27350	29969
8267	26189	3302

Sorgente 6.1: File RNS che configura il numero di esecuzioni (i. e., 5) e i semi numerici del modello di simulazione CORSIM relativo alla rete stradale del dataset #2.

Di seguito si descrivono le caratteristiche del dataset:

- tempo totale pari a 86400 secondi (i. e., 24 ore)
- suddivisione in 864 elementi (i. e., file)

³⁶ CORSIM supporta l'utilizzo di file Random Number Seed (RNS) per la configurazione di alcuni aspetti della simulazione: numero di esecuzione e semi numerici da utilizzare per la generazione stocastica di eventi.

³⁷ Il *seed* è un numero utilizzato per inizializzare un generatore di numeri pseudo-casuali. Condividendo tale numero e utilizzando lo stesso generatore di numeri pseudo-casuali utilizzato da CORSIM è possibile riprodurre la stessa identica simulazione su altre macchine.

- ogni elemento include 100 secondi di traiettorie di ogni sensore della succitata rete stradale
- ogni elemento è stato ottimizzato, cioè sottoposto all'eliminazione delle righe in cui non si verifica alcuna transizione di stato dei sensori
- la colonna `tp` è stata rimossa e sostituita con la colonna `class` in base alla corrispondenza riportata nella tabella 6.7 a pagina 77 (gli identificatori delle classi variano come detto per i profili di traffico relativi al sabato e alla domenica)
- l'intestazione di ogni elemento è:
`time, class, D5096, D341, D5098, D5099, D211, D212, D241, D242, D5097, D421, D5095, D541, D5094, D641, D891, D892, D981, D982, D911, D912, D191, D192, D1290, D1682, D1681, D1687, D1110, D1091, D1092, D1480, D781, D782.`

Si osservi come tali caratteristiche sono indipendenti dal profilo di traffico simulato. Le differenze tra i 3 diversi modelli di traffico sviluppati si sostanziano, chiaramente, nel diverso andamento dei flussi di traffico sulla rete stradale, il che porta a diversi dati di rilevazione da parte dei sensori; nel numero di esecuzioni della simulazione (i.e., 5 per i giorni lavorativi e 1 per il sabato e la domenica) e nei valori assunti dalla colonna relativa alla classe di traffico.

6.2.2 Risultati

...

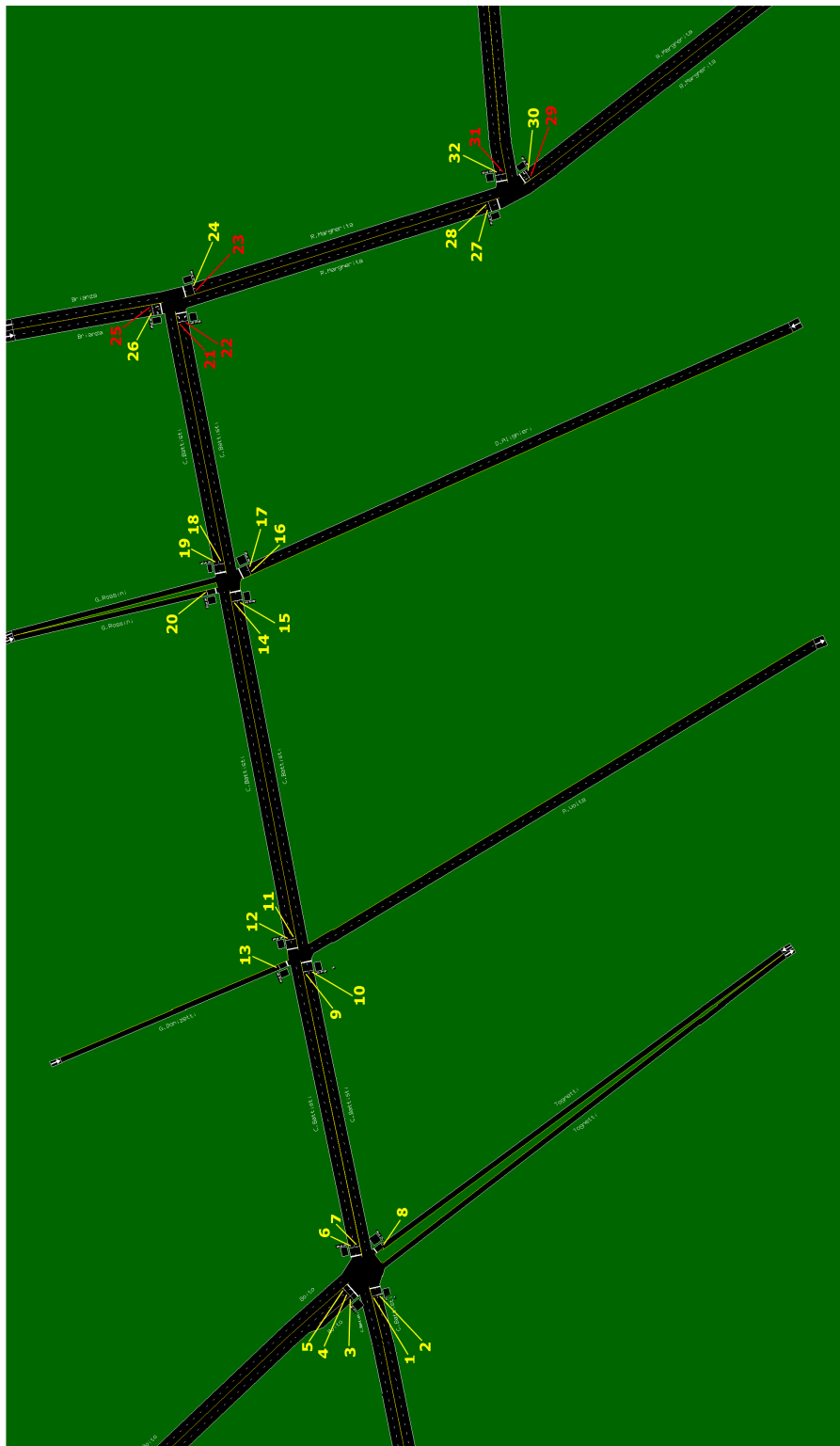


Figura 6.5: Visualizzazione del file TNO rappresentante la rete stradale da cui viene generato il dataset #2.

7 | CONCLUSIONI

...

A | GUIDE ALL'USO

A.1 UTILIZZO DEL PACKAGE CTBN

...

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

...

A.2.1 Sensors DLL

...

Installazione

...

Guida all'uso

...

A.2.2 Applicativi di supporto

...

ACRONIMI

API	Application Programming Interface	50
BN	Bayesian Network	1
BNC	Bayesian Network classifier	22
CIM	Conditional Intensity Matrix	10
CORSIM	Corridor microscopic simulation program	45
COM	Component Object Model	51
CPD	Conditional Probability Distribution	2
CPT	Conditional Probability Table	2
CSV	Comma Separated Values	60
CTBN	Continuos time Bayesian Network	1
CTBNC	Continuos time Bayesian Network classifier	21
CTNB	Continuos time Naive Bayes	21
CTNBC	Continuos time Naive Bayes classifier	22
CTTANB	Continuos time tree augumented Naive Bayes	21
CTTANBC	Continuos time tree augumented Naive Bayes classifier	22
DAG	Directed acyclic graph	1
DBN	Dynamic Bayesian Networks	33
DLL	Dynamic-link library	45
EM	Expectation Maximization	4
FRESIM	Freeway Simulator	46
FWHA	Federal Highway Administration	44
GUI	Graphical User Interface	44
IDE	Integrated Development Environment	44
IC	Inductive Causation	5
IM	Intensity Matrix	7
MAP	Maximum a posteriori	28
MCMC	Markov Chain Monte Carlo	5
MLE	Maximum Likelihood Estimation	18
NETSIM	Network Simulator	45
PV	Process Variable	11
RT	Record Type	45
RNS	Random Number Seed	79
RTE	Run-Time Extension	45
TAN	Tree Augumented Naive Bayes	22
TNO	TRAFED Native Object	45
TRAF	Traffic File	45
TRAFED	TRAF Editor	46
TRAFVU	TRAF Visualization Utility	46

TRF	Traffic File	44
TShell	TSIS Shell	46
TSIS	Traffic Software Integrated System	43
VBScript	Microsoft's Visual Basic Scripting Edition	46

INDICE ANALITICO

A

aciclicità 33, 36, 37
 apprendimento 16, 21
 apprendimento strutturale 37, 39

C

classificatore 21
 classificazione 2, 28
 conteggi immaginari 19, 20
 cross-validation 66

D

dati completi 28, 33, 66, 71
 dati di addestramento 33
 dati multinomiali 16
 Dirichlet 18, 36

E

esponenziale 8, 16–18

G

Gamma 18
 grafo 33, 34, 36
 greedy 37

H

hill climbing 33, 37–39

I

indipendenza globale 19
 indipendenza locale 19
 inferenza 21, 31

K

k-fold 66
 k-learn 36

L

likelihood marginale 34–36
 likelihood temporale 28, 29

M

maximum a posteriori 28
 multinomiale 17, 18

N

non supervisionata 21
 NP-completo 33, 36

O

onda quadra 43, 63
 ottimizzazione 33, 34, 36

P

parameter independence 34, 35
 parameter modularity 34
 parametri 33
 parametrizzazione mista 8, 9, 32
 parametrizzazione pura 8
 penalità del grafo 34
 polinomiale 33, 37
 priori coniugata 18, 19, 35
 pseudo-conteggi 19
 punteggio 33, 34, 37

R

regola di Bayes 28
 regolarizzazione bayesiana 16
 ricerca euristica 33

S

score bayesiano 33, 34, 36
 scoring 34, 36, 39
 segmenti temporali 12
 simulated annealing 38
 smoothing 19
 spire 43
 statistiche sufficienti 13–15, 18, 19
 stima bayesiana 18
 stima dei parametri 17, 18
 stime esatte 18
 structure modularity 34, 36
 struttura 36, 37
 strutturale 33
 supervisionata 21

T

tabu search 38
 time interval 69
 time period 69, 77
 training set 6, 26, 33, 34

V

valori attesi 20
 vettore aleatorio 29, 30

BIBLIOGRAFIA

Angulo, Eusebio, Francisco P Romero, Ricardo García, Jesús Serrano-Guerrero e José A Olivas

- 2011 «An adaptive approach to enhanced traffic signal optimization by using soft-computing techniques», *Expert Systems with Applications*, 38, 3, p. 2235-2247. (Citato a p. 42.)

Certet

- 2004 «Sei felice di essere un'automobilista». (Citato a p. 42.)

Chickering, David Maxwell

- 1994 *Learning Bayesian networks is NP-hard*, rapp. tecn., Microsoft Research. (Citato alle p. 33, 36.)
- 2013 «A Transformational Characterization of Equivalent Bayesian Network Structures», *CoRR*, p. 87-98, <http://arxiv.org/abs/1302.4938>. (Citato a p. 6.)

Chickering, David Maxwell, David Heckerman e Christopher Meek

- 2004 «Large-sample learning of Bayesian networks is NP-hard», ... *Journal of Machine Learning* ..., 5, p. 1287-1330, <http://dl.acm.org/citation.cfm?id=1044703>. (Citato a p. 33.)

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. 4.)

Duda, R. O. e P. E. Hart

- 1973 *Pattern Classification and Scene Analysis*, John Willey & Sons, New York. (Citato a p. 21.)

Felici, G, G Rinaldi, A Sforza e K Truemper

- 2006 «A logic programming based approach for on-line traffic control», *Transportation Research Part C: Emerging Technologies*, 14, 3 (giu. 2006), p. 175-189, ISSN: 0968-090X, DOI: 10.1016/j.trc.2006.05.007, <http://linkinghub.elsevier.com/retrieve/pii/S0968090X06000374>. (Citato a p. 42.)

Friedman, N, D Geiger e M Goldszmidt

- 1997 «Bayesian network classifiers», *Machine learning*, 163, p. 131-163, <http://link.springer.com/article/10.1023/A:1007465528199>. (Citato alle p. 21, 22, 29.)

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. 5.)

Gershenson, Carlos

- 2008 «Self-Organizing traffic lights», *Arxiv preprint nlin/0411066*, <http://arxiv.org/abs/nlin.A0/0411066>. (Citato a p. 42.)

Gihman, Iosif I. e Anatolij V. Skorohod

- 1973 *The theory of stochastic processes II*, New York: Springer-Verlag. (Citato a p. 9.)

Gilks, WR, S Richardson e DJ Spiegelhalter

- 1996 «Markov chain Monte Carlo in practice». (Citato a p. 5.)

Gualtieri, M, L Rigamonti, V Galeotti e MC Camatini

- 2005 «Toxicity of tire debris extracts on human lung cell line A549», *Toxicology in Vitro*, 19, 7 (ott. 2005), p. 1001-1008, ISSN: 0887-2333, DOI: 10.1016/j.tiv.2005.06.038, <http://hdl.handle.net/10281/15238>. (Citato a p. 42.)

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 alle p. 5, 18, 19.)

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 alle p. 6, 19.)

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

- 2011 *Bayesian Artificial Intelligence*, Chapman & Hall / CRC Computer Science and Data Analysis, CRC Press, ISBN: 9781439815915. (Citato alle p. 1, 3.)

Langley, Pat, Wayne Iba e Kevin Thompson

- 1992 «An Analysis of Bayesian Classifiers», in *AAAI*, p. 223-228, ISBN: 0-262-51063-4. (Citato alle p. 21, 23.)

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. 6, 7.)

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. 5.)

Mantecca, P, M Gualtieri, M Andrioletti, R Bacchetta, C Vismara, G Vailati e MC Camatini

- 2007 «Tire debris organic extract affects Xenopus development», *Environment international*, 33, 5, p. 642-648, ISSN: 0160-4120, <http://hdl.handle.net/10281/22470>. (Citato a p. 42.)

Nodelman, Uri D.

- 2007 *Continuous Time Bayesian Networks*, tesi di dott., Stanford University. (Citato alle p. 9, 11, 12, 16, 18.)

Nodelman, Uri D., 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. 15, 33, 37.)

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. 7.)

Papageorgiou, M., C. Diakaki, V. Dinopoulou, A. Kotsialos e Y. Wang

- 2005 «Review of road traffic control strategies», *Proceedings of the IEEE*, 91, 12 (dic. 2005), p. 2043-2067, ISSN: 0018-9219, DOI: 10.1109/JPROC.2003.819606, http://ieeexplore.ieee.org/xpls/abs%5C_all.jsp?arnumber=1246386. (Citato a p. 42.)

Park, Byungkyu e CJ Messer

- 1998 «A Genetic Algorithm-Based Signal Optimization Program for Oversaturated Intersections», *Texas Transportation Institute*, p. 1-8. (Citato a p. 42.)

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. 5.)

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. 2, 3, 37, 38.)

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. 5.)

Steck, Harald and Jaakkola, Tommi S

- 2002 «On the Dirichlet Prior and Bayesian Regularization», *Advances in Neural Information Processing Systems*, September, p. 713-720. (Citato a p. 19.)

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, 12, 22, 25, 28, 31, 33.)

Thorpe, T.L. e C.W. Anderson

- 1996 «Traffic light control using sarsa with three state representations», in *IBM Corporation, Citeseer*, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.5406>. (Citato a p. 42.)

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. 5.)

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, settembre 2013

Leonardo Di Donato