

Laurea Specialistica in Informatica
a.a. 2007-2008

Interazione Uomo-Macchina II:

Interfacce Intelligenti

Fiorella de Rosis

Laboratori di
Irene Mazzotta e Nicole Novielli

1

Introduzione

Unità 1: Ragionamento logico:

- Formalizzazione
- Risoluzione

Unità 2: Generazione di linguaggio naturale

- Teorie
- Metodi

***Unità 3:* Ragionamento incerto**

- Reti Causali Probabilistiche
- Reti dinamiche

Unità 4: Modelli di Utente

- Modelli logici
- Modelli con incertezza

Unità 5: Comprensione del linguaggio naturale

Unità 6: Simulazione di dialoghi

- Modelli basati su ATN
- Modelli basati su IS

Unità 7: Affective Computing

Programma
del Corso

2

Tratteremo questi argomenti

1. Una premessa sul concetto di probabilità
2. Cosa sono le reti causali probabilistiche (RCP)
3. Come si possono costruire:
 - In modo totalmente manuale
 - Mediante algoritmi di apprendimento
4. Come si possono utilizzare: ragionamento 'diagnostico' e 'prognostico'
5. Quando una RCP diventa 'dinamica'

3

Consideriamo di nuovo l'esempio Es 2.5

*"Mangiare ad orari fissi **aiuta ad** evitare di saltare i pasti ..."*

Nella Unità 2, avevamo formalizzato questa frase come:

$\forall x \text{ (Person}(x) \wedge \text{EatAtFixedTime}(x)) \rightarrow \text{AvoidJumpMeal}(x)$

Ma la formalizzazione con linguaggio logico radicalizza l'affermazione

(non necessariamente si evita di saltare i pasti se si mangia in modo regolare).

E' necessario, in questo caso, 'rilassare' l'implicazione introducendo un elemento d'incertezza.

$\forall x \text{ (Person}(x) \wedge \text{EatAtFixedTime}(x)) \rightarrow? \text{AvoidJumpMeal}(x)$

L'implicazione avrà, in questo modo, un *grado di forza*.

4

Consideriamo ora l'esempio Es 2.4

Nella Unità 3, abbiamo definito come 'certi' i dati:

$\text{Implies}(a,g), \text{Likes}(x,g), \text{CanDo}(x,a);$

mentre potrebbero avere un '*grado di certezza*':

'è ampiamente dimostrato che', 'mi sembra che',

'probabilmente', ... ecc

così come l'implicazione:

$\forall x \forall a \forall g$

$((\text{Implies}(a,g) \wedge \text{Likes}(x,g) \wedge \text{CanDo}(x,a)) \rightarrow ?(\text{ShouldDo}(x,a)))$

Vediamo come si può formalizzare la conoscenza e come si può ragionare in condizioni di incertezza.

*Sceghlieremo, in particolare, le misure di incertezza che si basano sulla **teoria della probabilità**.*

Cominciamo quindi con qualche chiarimento sul concetto di probabilità.

5

1. Premessa sui diversi concetti di probabilità

1. Definizione 'classica' (Laplace, 1820)

*Probabilità come rapporto fra il numero di casi **favorevoli** all'evento di cui si vuole misurare la probabilità e il **n** totale di casi **possibili**.*

Es: probabilità che esca un picche in un mazzo di 52 carte è .25 (1/4)

2. Idea 'frequentista' (Von Mises, 1928)

Probabilità come frequenza-limite del verificarsi di un

evento E: $P(E) = \lim_{n \rightarrow \infty} S_n(E) / n$

Al crescere di n (ripetizioni dell'esperimento), le frequenze si stabilizzano fornendo una stima sempre meno approssimata del valore di probabilità.

6

Limiti delle concezioni 'classiche'

Le due teorie viste finora sostengono che affermazioni come

"Probabilmente, la Roma non vincerà il prossimo campionato"

non rientrano nel dominio dei problemi trattati dalla teoria della probabilità.

Infatti: non si può calcolare il n. di casi favorevoli su quelli possibili; e non si può effettuare un esperimento ripetuto nelle stesse condizioni.

Tuttavia, il mondo delle scommesse assegna comunemente un valore a queste affermazioni.

Questo valore viene stabilito sulla base del *grado di fiducia* (degree of belief) sul verificarsi dell'evento in gioco:

metodo di misura tipicamente soggettivo!!

7

Un'idea (relativamente) nuova di probabilità

3. Probabilità soggettiva o 'neo-bayesiana' (De Finetti, 1960)

*si basa su una stima soggettiva di probabilità di un evento come 'somma che una persona è disponibile a **scommettere** sull'evento'. Due ipotesi:*

- **scambiabilità**: i due scommettitori sono disponibili a scambiarsi i termini della scommessa;
- **'non desiderio di perdere'**: nessuno dei due è disponibile a stipulare scommesse che diano certamente una perdita.

*Le due ipotesi garantiscono che l'insieme di previsioni sia **'coerente'** (tra le combinazioni di scommesse che ci s'impegna ad accettare, non ve ne sono che diano guadagni tutti uniformemente negativi).*

<http://www.answers.com/topic/bruno-de-finetti>

<http://www.controluce.it/giornali/a13n09/20-cultura-brunodefinetti3.htm>

8

Vediamo un esempio

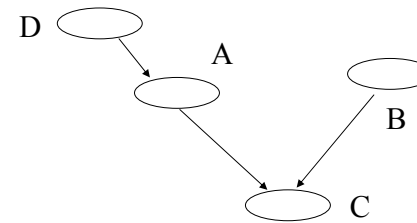
Qual è la probabilità che il PdL vinca le elezioni?
 Io penso (credo, sono convinta...) che sia = .48
 Questo significa che sono disponibile a pagare 48 euro, accettando di avere 100 euro se il PdL perde e 0 se vince.
 Ma significa anche che sarei disponibile, per coerenza (o che ci sia qualcun altro disponibile) a pagare 52 euro, accettando di avere 100 euro se il PdL vince e 0 se perde.
 Complessivamente, avremo raccolto 100 euro.

Quali sono i guadagni?

Se il PdL vince, chi aveva pagato 52 euro ne guadagna 48; se perde, chi aveva pagato 48 euro ne guadagna 52.

Considereremo, fra i vari metodi proposti per il ragionamento incerto, le *Reti Causali Probabilistiche*, che sono oggi il formalismo prevalente nel campo delle interfacce intelligenti. Come vedremo, le RCP possono essere costruite sulla base dell'idea 'soggettivista' o 'frequentista' di probabilità

2. Cosa sono le reti causali probabilistiche (RCP): cominciamo con un esempio semplice



In questa RCP sono rappresentate le variabili A,B,C,D. Queste possono essere variabili booleane, oppure a più valori.

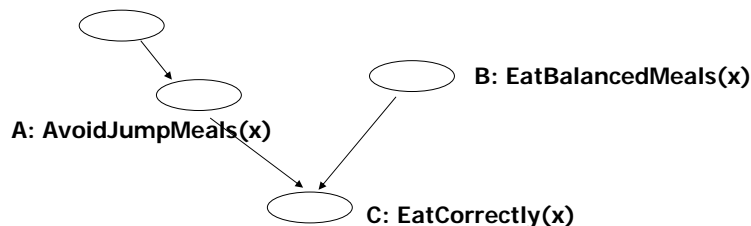
L'arco $D \rightarrow A$ rappresenta una *relazione causale* fra D ed A.

I parametri associati a questo arco rappresentano la '*forza*' di questa relazione.

Gli archi $A \rightarrow C$ e $B \rightarrow C$ rappresentano l'ipotesi che *sia A che B siano causa di C*, anche se (eventualmente) con '*forze*' differenti.

Cosa potrebbe rappresentare la RCP che abbiamo appena visto

D: EatAtFixedTime(x)



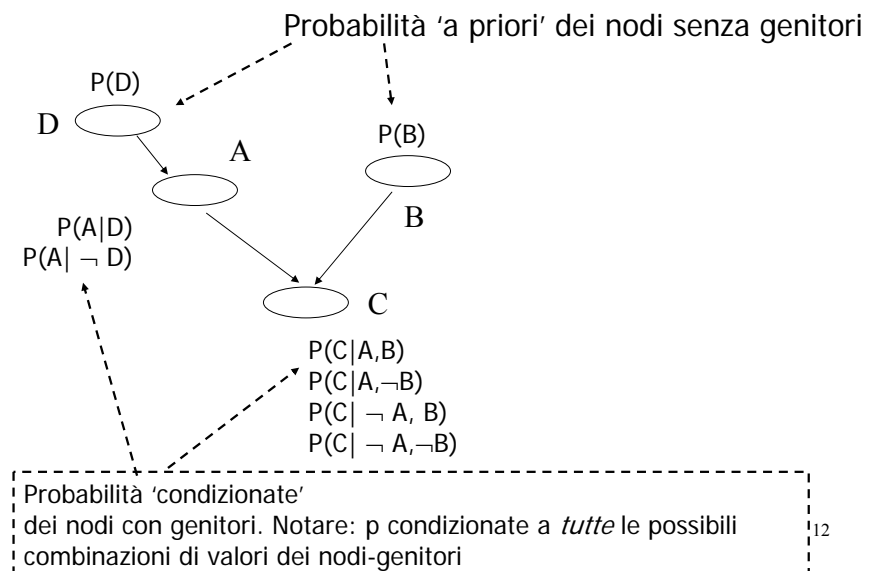
"Mangiare ad orari fissi aiuta ad evitare di saltare i pasti ..."

"Evitare di saltare i pasti e mangiare pasti bilanciati contribuiscono a mangiare correttamente"

In questo esempio, A,B,C,D sono variabili booleane

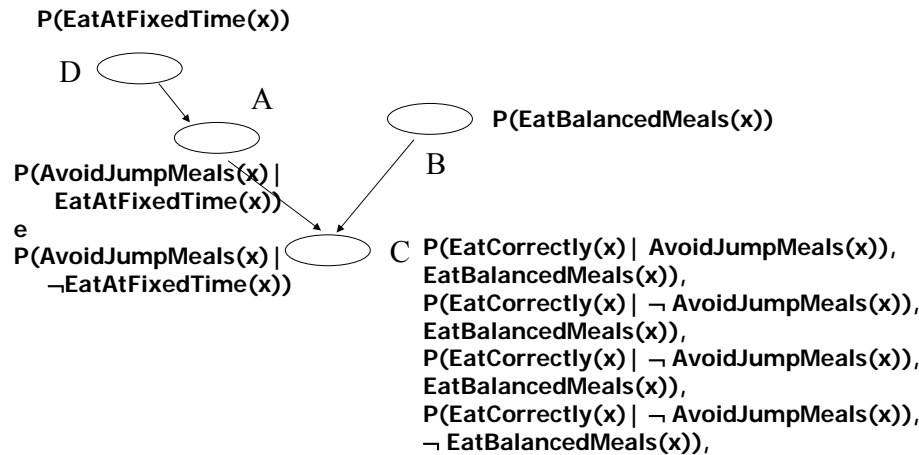
11

Quali parametri occorre associare ad una RCP



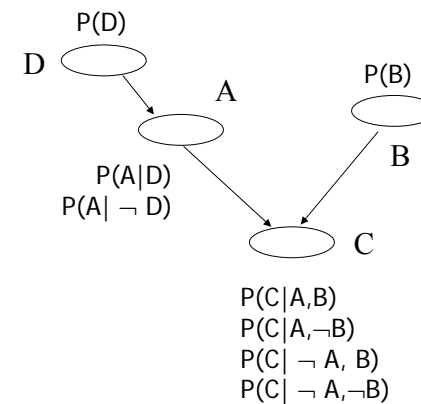
12

Nel nostro esempio



13

Ipotesi di indipendenza condizionale



Il grafo rappresenta ipotesi di 'indipendenza condizionale' tra le variabili associate ai suoi nodi:

A è indipendente da B e da C, condizionatamente a D:
 $P(A|B,C,D) = P(A|D)$

C è indipendente da D, condizionatamente ad A:
 $P(C|A,B,D) = P(C|A,B)$

14

Ora possiamo generalizzare la definizione di RCP

Grafi orientati senza cicli i cui nodi rappresentano variabili a più valori.

Gli archi rappresentano l'esistenza di *influenze causali* fra le variabili che collegano.

La forza di queste influenze è misurata in termini di *probabilità condizionate*.

Se il grafo contiene le variabili $x_1, x_2, \dots, x_i, \dots, x_n$,
 e G_i è l'insieme dei *genitori* di x_i ,

una quantificazione completa e consistente delle relazioni fra le variabili può essere ottenuta specificando,

$\forall x_i, \text{ la } P(x_i | G_i).$

15

Fasi nella Costruzione di una RCP

1. Definire la struttura del grafo:
 - semantica dei nodi
 - semantica delle relazioni
 - attenzione alle ipotesi di indipendenza condizionale!
2. Assegnare i parametri

16

Conseguenze dell'ipotesi di indipendenza condizionale

La distribuzione congiunta di probabilità (o 'probabilità marginale'):

$$P(x_1, x_2, \dots, x_i, \dots, x_n)$$

potrà essere approssimata con il prodotto delle probabilità condizionate:

$$P(x_1, x_2, \dots, x_i, \dots, x_n) = \prod_i P(x_i | G_i).$$

Nell'esempio precedente:

$$P(A, B, C, D) = P(C | A, B, D) * P(A | B, D) * P(B | D) * P(D) = \\ P(C | A, B) * P(A | D) * P(B) * P(D)$$

E quindi, la probabilità marginale può essere calcolata in funzione dei parametri inseriti nella rete.

17

Probabilità 'a priori' delle variabili

Per probabilità a priori di una variabile s'intende il valore di prob di questa variabile in assenza di 'evidenze', e cioè di 'fatti noti'.

Le "probabilità a priori" di ciascuna delle variabili nella rete possono essere calcolate dai parametri inseriti nella rete stessa.

Nell'esempio precedente:

$$P(A) = P(A, D) + P(A, \neg D) =$$

$$P(A | D) * P(D) + P(A | \neg D) * P(\neg D)$$

... e tutti questi sono parametri di input...

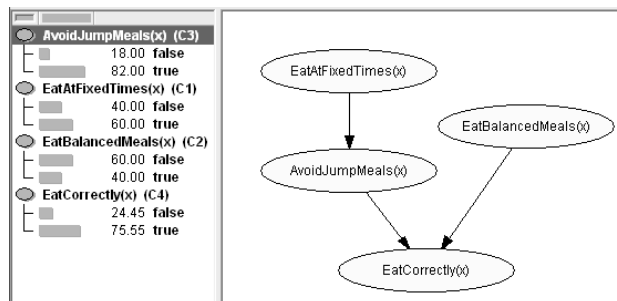
... analogamente per C.

Ma possono essere calcolate anche a partire dalle probabilità marginali:

$$P(A) = P(A, B, C, D) + P(A, \neg B, C, D) + P(A, B, \neg C, D) + \\ P(A, B, C, \neg D)$$

18

Nel nostro esempio (simulato con Hugin)



$$P(\text{AvoidJumpMeals}) = \\ P(\text{AvoidJumpMeals} | \text{EatAtFixedTimes}) * P(\text{EatAtFixedTimes}) + \\ P(\text{AvoidJumpMeals} | \neg \text{EatAtFixedTimes}) * P(\neg \text{EatAtFixedTimes}) = \\ = .9 * .6 + .7 * .4 = .82$$

19

Quindi...

Quando assegno i vari parametri ad una RCP, implicitamente ho definito:

- Le probabilità a priori di tutti i nodi
- La probabilità marginale della rete.

... a cosa mi serve questa distribuzione congiunta????

Lo vedremo fra un po'...

20

Come assegnare i parametri alla RCP?

Due metodi, che riflettono l'idea soggettivista e l'idea frequentista:

- Assegnare le prob a priori e condizionate sulla base della *conoscenza soggettiva del dominio*, e quindi del proprio *'grado di fiducia'* sull'associazione fra le variabili;
- 'Calcolare' questi parametri da una *misura di frequenze su un set di dati*, applicando metodi di apprendimento.

Nota:

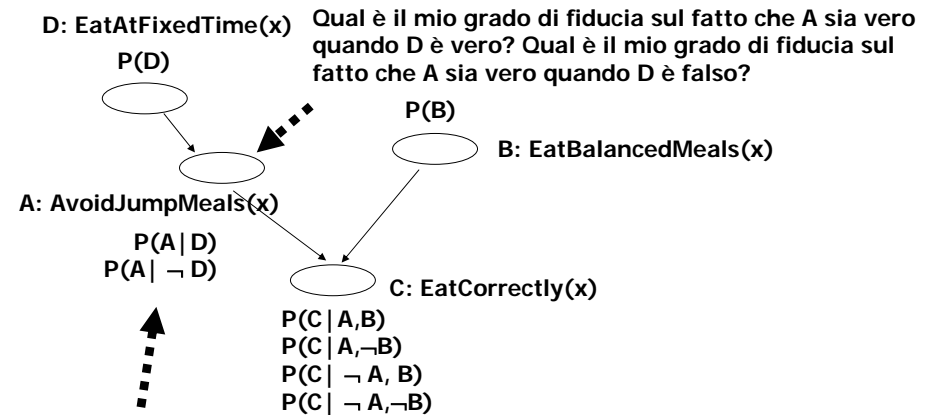
Il metodo a. richiede l'applicazione di procedure per la verifica della *coerenza* delle valutazioni.

Il metodo b. richiede la realizzazione di un *esperimento di raccolta dati* su un numero sufficientemente elevato di casi, rappresentativi della popolazione a cui la RCP dovrà poi essere applicata.

I due metodi possono essere integrati con una procedura di apprendimento interattivo della RCP.

21

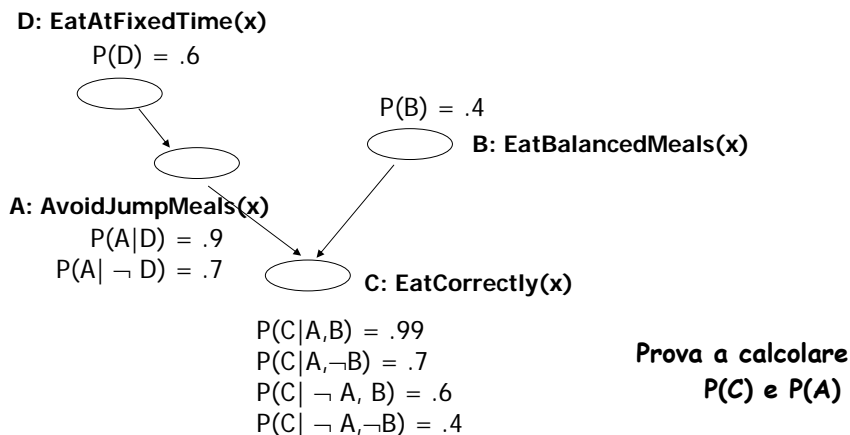
3.1 Costruzione manuale delle RCP: parametri come 'degree of belief'



Quanto è probabile che si eviti di saltare i pasti se si mangia ad orari fissi? Ad es, .9
Quanto è probabile che questo avvenga se *non* si mangia ad orari fissi? Ad es, .7

22

... esempi di valori...



23

Un algoritmo per la propagazione di 'evidenze' nelle RCP

(D. Spiegelhalter)

Cos'è una *'evidenza'*:

è la conoscenza del valore assunto da una variabile x_i nella RCP.

Nel nostro esempio precedente:

"Maria fa una dieta bilanciata", oppure *"Maria mangia correttamente"*.

L'evidenza può riguardare *uno qualsiasi dei nodi* della RCP, non necessariamente i nodi-radice.

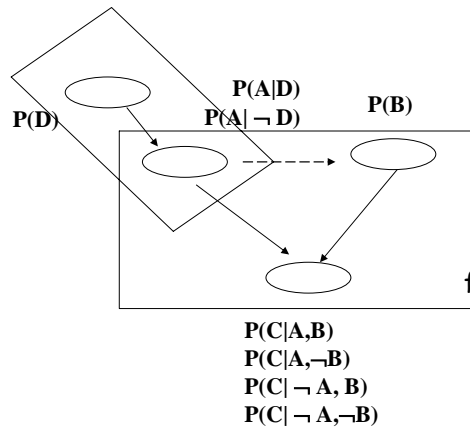
Propagare le evidenze significa calcolare quali sono gli effetti di questa conoscenza sui valori di probabilità delle altre variabili, che chiameremo *"probabilità a posteriori"*, che denoteremo con $P^*(x_j)$.

24

Outline dell'Algoritmo di Spiegelhalter

Il calcolo delle probabilità a priori e a posteriori dei nodi si può effettuare con un algoritmo di:

- moralizzazione del grafo,
- decomposizione del grafo in 'cricche'
- calcolo delle prob 'marginali' delle cricche.



Le prob a priori e posteriori di ogni nodo sono quindi calcolabili in funzione delle prob marginali delle cricche e delle loro intersezioni. Le prob marginali del grafo sono funzione del rapporto fra:

- prodotto delle prob delle cricche e
- prodotto delle prob delle loro intersezioni

Vediamo meglio:

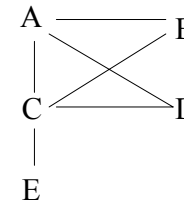
Cosa è una 'cricca'

Dato un grafo non orientato G con un insieme V di nodi.

Un sottoinsieme W di V si dice *cricca* di G se W è un *insieme completo massimo*, cioè

un grafo in cui tutte le coppie di nodi sono adiacenti e non c'è un altro insieme completo che lo contiene.

Esempio:



ABC e CDE non sono cricche

ABCD è una cricca

CE è una cricca

C è l'intersezione fra le cricche

ABCD e CE

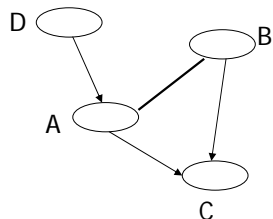
La probabilità marginale di G può essere espressa in funzione prob marginali delle sue cricche, in un grafo non orientato equivalente

Moralizzazione di un grafo

La 'moralizzazione' di un grafo consiste nel trasformare un grafo orientato in un grafo *non* orientato *equivalente* introducendo nuovi archi fra i nodi-genitori dello stesso nodo e

lasciando i parametri invariati.

Esempio: moralizziamo il grafo del nostro esempio:



Probabilità a priori dei nodi

Si possono calcolare, sul grafo non orientato 'moralizzato',

in funzione delle

probabilità delle cricche.

Ad esempio: (A,D) è una cricca.

$$P(A) = P(A,D) + P(A,\neg D)$$

Anche (A,B,C) è una cricca

$$P(C) = P(A,B,C) + P(\neg A,B,C) + P(A,\neg B,C) + P(\neg A,\neg B,C)$$

... ecc

Anche le 'probabilità marginali' si possono calcolare in funzione delle probabilità delle cricche

$$P(A,B,C,D) = P(C|A,B,D) * P(A|B,D) * P(B|D) * P(D)$$

Semplificando, come abbiamo visto:

$$P(A,B,C,D) = P(C|A,B) * P(A|D) * P(B) * P(D)$$

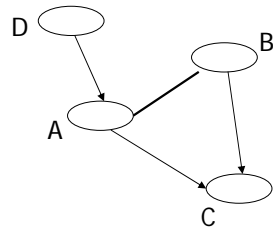
Ma

$$P(C|A,B) = P(A,B,C) / P(A,B) = P(A,B,C) / (P(A) * P(B))$$

$$P(A|D) * P(D) = P(A,D)$$

Quindi

$$P(A,B,C,D) = (P(A,B,C) * P(A,D)) / P(A)$$



Probabilità marginale =
rapporto fra:
prodotto delle probabilità
delle cricche
e
probabilità
delle loro intersezioni

29

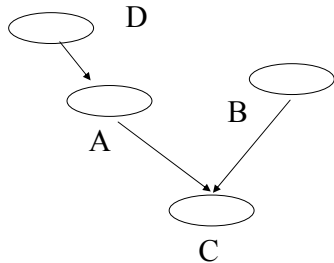
Vediamo ora come calcolare le probabilità a posteriori, dopo inserimento di una o più evidenze

**Data una 'evidenza' (valore assunto da alcune delle variabili rappresentate nel grafo),
Valutare come questa evidenza influisce
sulla probabilità delle altre variabili
rappresentate nel grafo.**

Notare: se le variabili sono booleane, le evidenze corrisponderanno a 'valori di verità'.

30

Aggiornamento della probabilità delle cricche



Supponiamo di avere un'evidenza sul nodo C;
Ad esempio, C è 'vero': $P(C) = 1$.

Vogliamo valutare come questa conoscenza influisce sulla probabilità dell'evento associato al nodo A: $P^*(A) = P(A|C)$.

A è l'intersezione fra le due cricche AD e ABC.

$$\begin{aligned} \text{Consideriamo la prob a posteriori della cricca AD: } P^*(A,D) &= P(A,D|C) = P(A,D,C) / P(C) = \\ &= P(D|A,C) * P(A|C) = \\ &= P(D|A) * P(A|C) = \\ &= P(A,D) / P(A) * P(A|C) = \\ &= P(A,D) * (P^*(A) / P(A)) \end{aligned}$$

Cioè:

La probabilità a posteriori della cricca AD è =
alla sua probabilità a priori, per il rapporto fra prob a posteriori e a priori
dell'intersezione fra le due cricche AD e ABC.

31

Riassumiamo ...

In un grafo 'moralizzato':

- la **probabilità a priori di una cricca** è funzione dei parametri inseriti nella rete:

$$P(A,B,C) = P(C|A,B) * P(A) * P(B) = P(C|A,B) * P(A|D) * P(D) * P(B)$$

- la **probabilità a posteriori di una cricca** è il prodotto della sua probabilità a priori, per il rapporto fra prob a posteriori e a priori dell'intersezione con una cricca 'adiacente':

$$P^*(A,B,C) = P(A,B,C) * (P^*(A) / P(A))$$

- la **probabilità marginale del grafo** è il rapporto fra prodotto delle probabilità marginali delle cricche e probabilità della loro intersezione con una cricca 'adiacente':

$$P(A,B,C,D) = (P(A,B,C) * P(A,D)) / P(A)$$

- la **probabilità a priori dei nodi** è la somma delle probabilità a priori delle cricche:

$$P(C) = P(A,B,C) + P(\neg A,B,C) + P(A,\neg B,C) + P(\neg A,\neg B,C)$$

o delle probabilità marginali a priori della rete:

$$P(C) = P(A,B,C,D) + P(\neg A,B,C,D) + P(A,\neg B,C,D) + P(\neg A,\neg B,C,D) \dots$$

Probabilità a posteriori dei nodi

Come per le probabilità a priori, anche le probabilità a posteriori dei nodi si possono calcolare dalla probabilità a posteriori delle cricche:

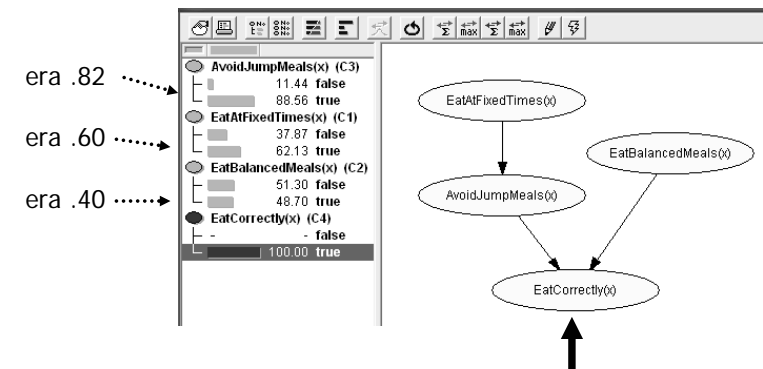
$$P^*(A) = P^*(A, B, C) + P^*(A, \neg B, C) + P^*(A, B, \neg C) + P^*(A, \neg B, \neg C)$$

$$P^*(C) = P^*(A, B, C) + P^*(\neg A, B, C) + P^*(A, \neg B, C) + P^*(\neg A, \neg B, C)$$

... ecc

33

Nel nostro esempio (sempre con Hugin)



So che l'individuo considerato (Maria) 'mangia correttamente':
 $P(\text{EatCorrectly}(x)) = 1$.

Qual è, ora la probabilità che x mangi ad orari fissi?
 Qual è la probabilità che segua una dieta corretta?

34

2.2. Costruzione di RCP: mediante algoritmi di apprendimento

Passi necessari

- Piano della raccolta dati (esperimento):
 - definizione della popolazione su cui effettuare l'esperimento
 - definizione delle variabili da 'osservare'
 - definizione dei metodi di codifica (var continue vs discrete e valori possibili)
- Codifica dei dati e verifica degli errori
- Discretizzazione delle variabili continue
- Calcolo delle prob a priori e condizionate
- Valutazione dell'*attendibilità* del modello
- Test preliminare del modello
- Eventuale revisione iterativa della struttura e dei parametri
- Test finale del modello in situazioni 'significative'

35

Definizione di una struttura semplificata

Se il database contiene n variabili, un belief network che corrisponde ad un grafo 'completo' (con n nodi) conterrà
 $n*(n-1)/2$ archi (con orientamento da determinare).

E' quindi necessario determinare *il BN con il numero minimo di archi che rappresenta in modo 'accettabile' il dominio considerato.*

Formulazione del problema.

Date:

n variabili, rappresentate ciascuna in un nodo di un grafo orientato G ;
 un dataset $D = \{ \langle v_1^1, \dots, v_n^1 \rangle, \dots, \langle v_1^h, \dots, v_n^h \rangle \}$ relativo all'osservazione dei valori delle n variabili su un numero h di 'casi'.

Trovare il modello M che meglio approssima D , e cioè che massimizza la $P(D | M)$.

Individuare il modello corrisponde a individuare:

- quali archi legano i nodi in G (cioè, la sua *struttura*) e
- quali *parametri* sono associati ai nodi di G (probabilità a priori delle radici, e condizionate per gli altri nodi).

36

Una misura di qualità del modello M:

Marginal Log Likelihood di una RCP:

$$P(D | M) = \prod_k P(v^k | M)$$

Gli algoritmi di apprendimento si propongono di trovare la **MLL minima**, dato un dominio descritto da un insieme di variabili e un set di dati che descrive il comportamento di queste variabili in un numero ‘appropriato’ di soggetti.

La MLL di una RCP è quindi uguale alla somma delle MLL dei suoi nodi.
Dipende dal numero di nodi nella RCP e (in misura molto minore) dalla sua struttura.

37

Outline dell'algoritmo

- \forall nodo della rete, si calcolano le correlazioni con gli altri nodi (candidati genitori), fino ad un numero massimo di genitori prefissato;
- per ogni combinazione (nodo considerato, genitori candidati), si calcola la MLL del nodo;
- si ordinano le combinazioni per valori decrescenti della MLL;
- si costruisce iterativamente la rete collegando i nodi con i genitori individuati;
- si calcola la MLL della rete.

38

Revisione dei risultati dell'apprendimento

L'algoritmo non tiene conto del significato delle variabili né dell'uso che verrà fatto del modello.

Quindi un modello ‘ottimale’ in termini di MLL può non essere *interpretabile*.

Questo limite implica che un apprendimento totalmente automatico del modello produce, in genere, risultati insoddisfacenti

e che la strategia da adottare per la costruzione di un buon modello debba essere basata, piuttosto, su una *integrazione di procedura manuale e automatica*, e su un *processo iterativo di costruzione del modello* (per ‘raffinamenti successivi’).

39

Confrontiamo i due metodi di costruzione di una RCP su un esempio

40

Riprendiamo l'esempio della persuasione (Es 2.4)

- Task model: le strategie persuasive di Walton**

$\forall x \forall a \forall g ((\text{Implies}(a,g) \wedge \text{Likes}(x,g) \wedge \text{CanDo}(x,a)) \rightarrow (\text{ShouldDo}(x,a)))$

$\forall x \forall a \forall g ((\text{Implies}(a,g) \wedge \neg \text{Likes}(x,g) \wedge \text{CanAvoid}(x,a)) \rightarrow \neg (\text{ShouldDo}(x,a)))$

- Domain model**

$\forall s \forall x (\text{Sport}(s) \rightarrow \text{Implies}(s, \text{InShape}))$

$\forall s \forall x (\text{Sport}(s) \rightarrow \text{Implies}(s, \text{GoodHealth}))$

$\forall s \forall x ((\text{Sport}(s) \wedge \text{Young}(x) \wedge \text{Healthy}(x)) \rightarrow \text{CanDo}(x,s))$

$\text{Sport}(R)$

- User model**

$\text{Young}(G)$

$\text{Healthy}(G)$

$\text{Likes}(G, \text{InShape})$

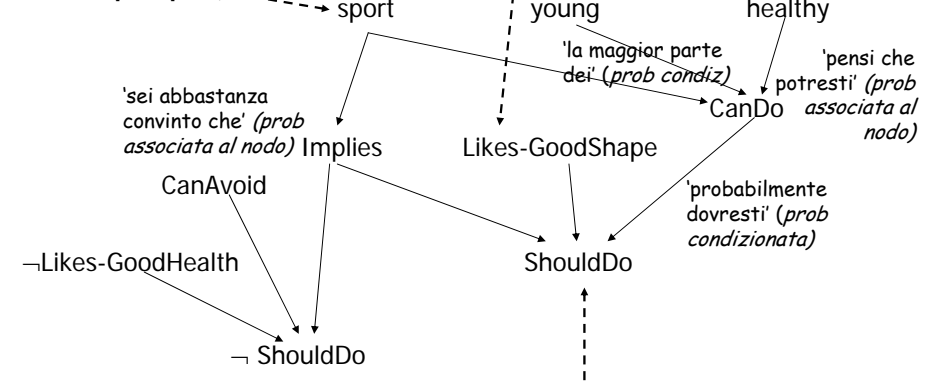
.....

41

Struttura del grafo

nel problema di persuasione rappresentato in modo *logico*

Qui inseriamo la 'conoscenza specifica' (evidenza su un particolare utente e uno o più sport)

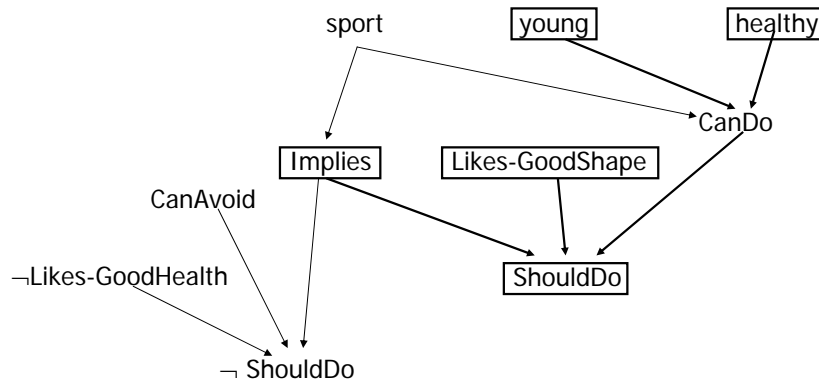


... e 'leggevamo' lo sport da suggerire

42

Struttura del grafo

nel problema di persuasione rappresentato in modo *logico*



... ricostruendo, all'indietro, il percorso seguito, per generare un messaggio di spiegazione

43

Come cambiano le cose, nella rappresentazione dello stesso problema in condizioni di incertezza

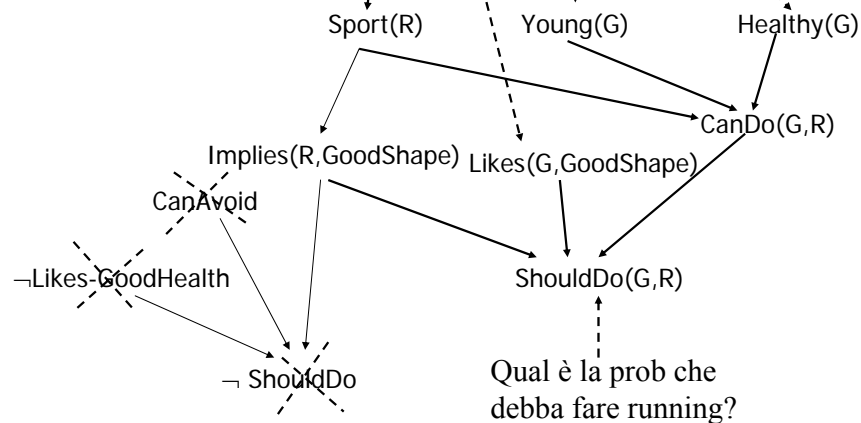
- Associa, ai nodi del grafo, delle variabili che rappresentano 'ground formulae', cioè formule istanziate.
- Sparisce la distinzione fra 'conoscenza generale' e 'conoscenza specifica'.

Ho perso in generalità, ma posso introdurre diversi elementi di incertezza

44

Ad esempio: ragiono su Giuseppe e Running

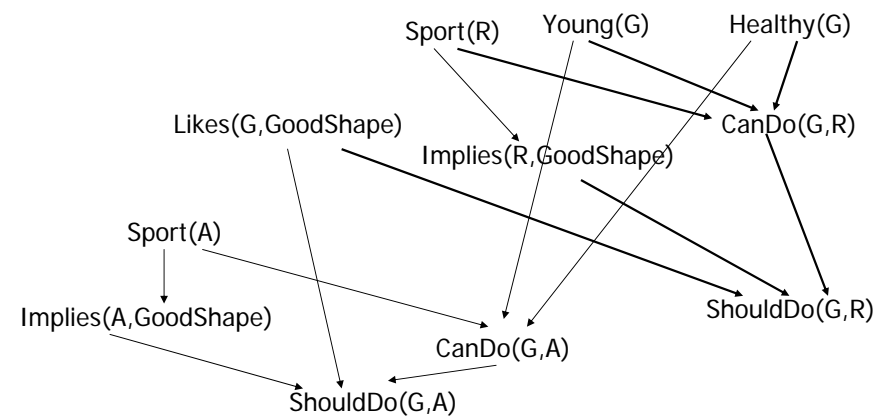
“Giuseppe è giovane e sano e ci tiene a stare in buona salute”



Posso semplificare il grafo

45

... oppure su Giuseppe, Running e AIKIDO



46

Come cambia l'algoritmo di ricerca del suggerimento più appropriato?

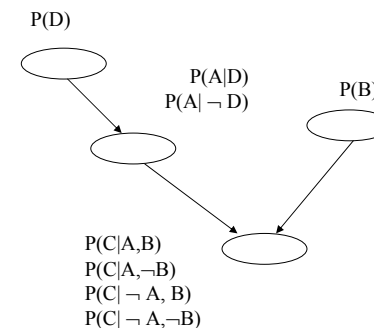
- Introduco nella rete l'evidenza (conoscenze che ho sul dominio e su Giuseppe)
- Propago
- Osservo i nodi *'ShouldDo(Giuseppe,s)'*
- Cerco quello con probabilità massima: questo corrisponde allo sport 'ottimale' per Giuseppe.

Che messaggi posso generare?

- *Quale sport* deve fare Giuseppe
- *Perché*: osservando i genitori di questo nodo e traducendo in linguaggio naturale sia il loro significato che il loro valore di probabilità

47

4. Come si possono utilizzare le RCP: ragionamento 'diagnostico' e 'prognostico'

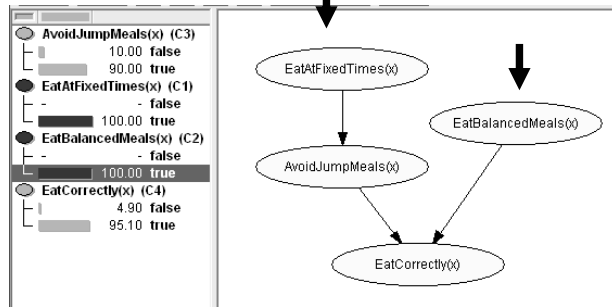


Si possono valutare le conseguenze, sull'intero grafo, di 'evidenza' acquisita su uno o più nodi:

- ragionando *dalle cause agli effetti* (rag 'prognostico'): noto il valore di D, valutare la prob di A,B,C
- ragionando *dagli effetti alle cause* (rag 'diagnostico'): noto il valore di C, valutare la prob di A, B, D (come abbiamo visto prima)

48

Un esempio di ragionamento prognostico



So che l'individuo considerato 'mangia ad orari fissi' e 'segue una dieta bilanciata':
 $P(\text{EatAtFixedTimes}(x) = 1)$; $P(\text{EatBalancedMeals}(x) = 1)$.
 Qual è, la probabilità che x mangi correttamente??

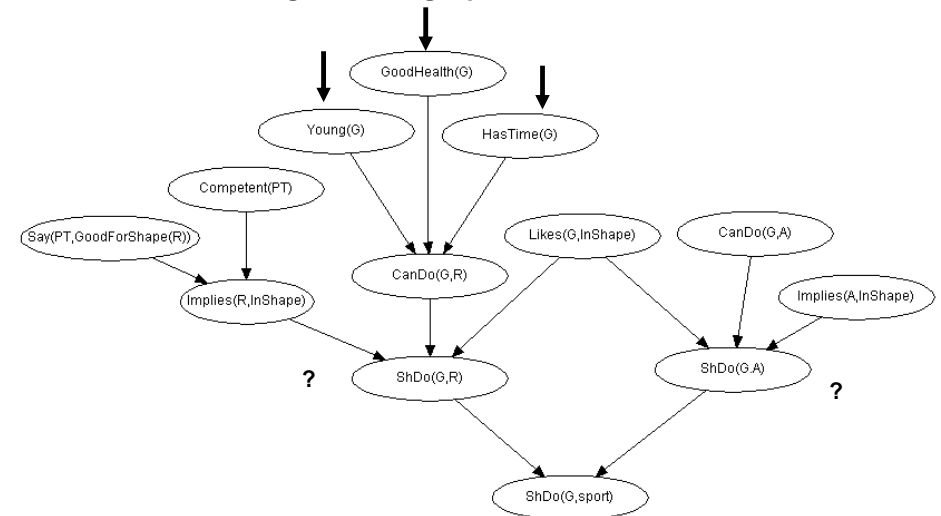
49

Come utilizzare le RCP in dialoghi persuasivi

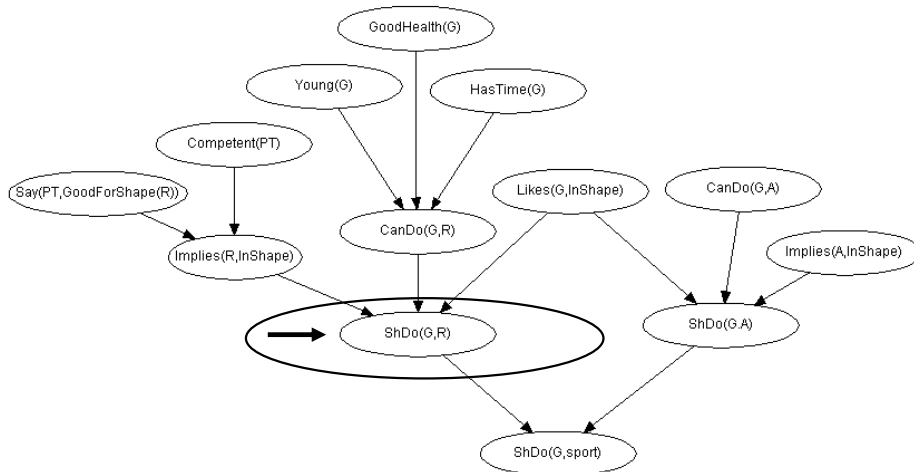
1. Inserisco evidenza sui nodi 'osservabili'.

Propago.

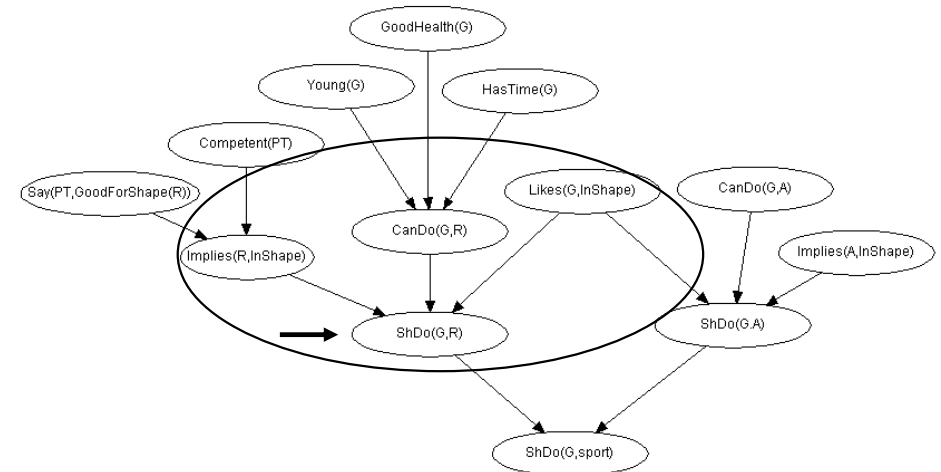
Scelgo la strategia persuasiva 'ottimale'



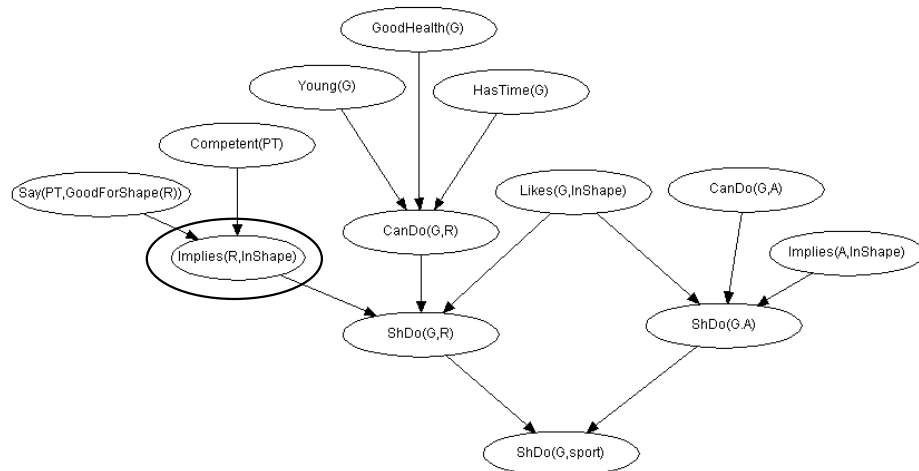
2. Genero il suggerimento:
 traduco il nodo ShDo in linguaggio naturale.
 Aspetto la reazione di G.



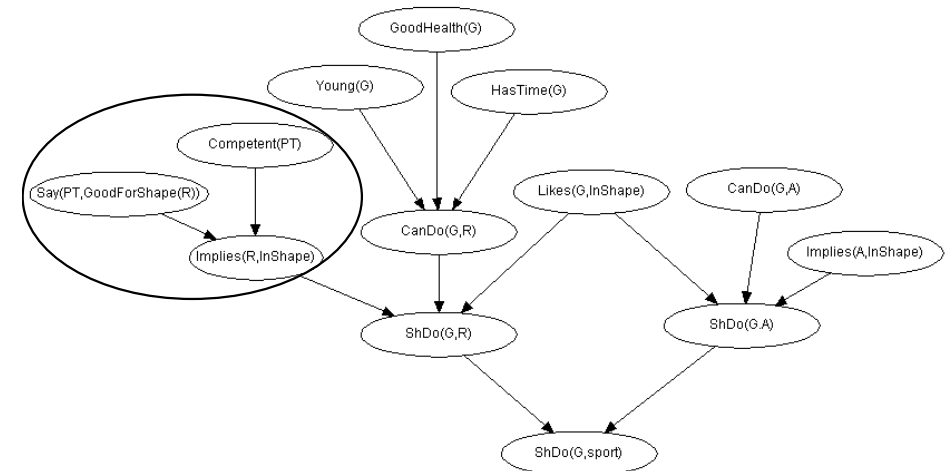
3. G chiede 'perché?'
 Genero una giustificazione:
 Osservo i genitori del nodo.
 Traduco ciascuno di essi in linguaggio naturale.
 Collego con dei connettivi linguistici



4. G chiede 'ma sei proprio sicuro che correre aiuti a stare in forma?'
 Rispondo traducendo il valore di probabilità del nodo in linguaggio naturale.

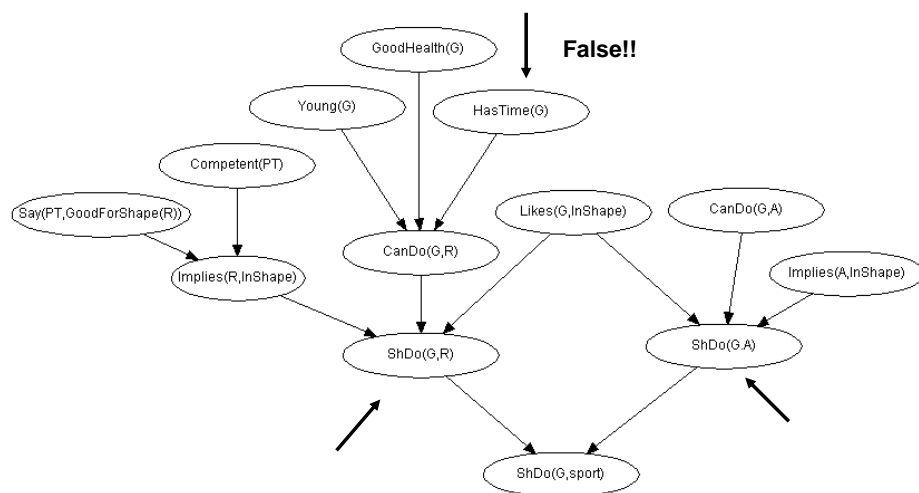


5. G chiede 'ma chi te l'ha detto?'
 Osservo i genitori del nodo e traduco in linguaggio naturale.



6. G obietta 'ma io non ho tempo di andare a correre!'
 Propago il nuovo dato nella rete.
 Osservo le conseguenze di questa nuova evidenza.
 Mi chiedo: devo cambiare strategia persuasiva?

... ecc...



5. Quando una RCP diventa 'dinamica'

Durante l'interazione, il modello dell'utente (in genere) cambia.
 Uno stesso dato può essere osservato più volte nel tempo.
 Il valore di alcune variabili in una certa fase dell'interazione può dipendere dal valore osservato in tempi precedenti:
 non necessariamente, però, è *identico* a questo valore.
 Può esserci un effetto di *decadimento* o di *accrescimento*, che deve essere rappresentato.

Esempio (affective computing):

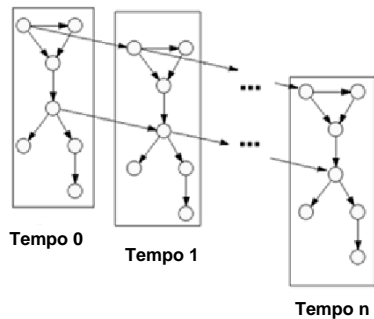
U fa una richiesta ad un Call Center;
 L'operatore non capisce;
 U 'si arrabbia' e ripete la sua richiesta;
 L'operatore rileva lo stato di 'rabbia' e cerca di placarlo;
 U è ancora arrabbiato, ma un po meno; ripete la sua richiesta;
 L'operatore capisce;
 La rabbia di U svanisce;

...

In questo caso, la variabile il cui andamento temporale deve essere rappresentato nel modello è l'emozione di 'rabbia'. L'osservabile può essere il tono di voce o il contenuto della sua 'mossa'

DBN: una definizione generale

Un modello dinamico è una sequenza di sottomodelli, ciascuno dei quali rappresenta lo stato del sistema in un certo istante di tempo



DBN = {V, E}, dove:

- $V(t_i)$ descrive la struttura del modello nell'istante di tempo t_i
- t_0 e t_n sono il primo e l'ultimo istante di tempo considerati

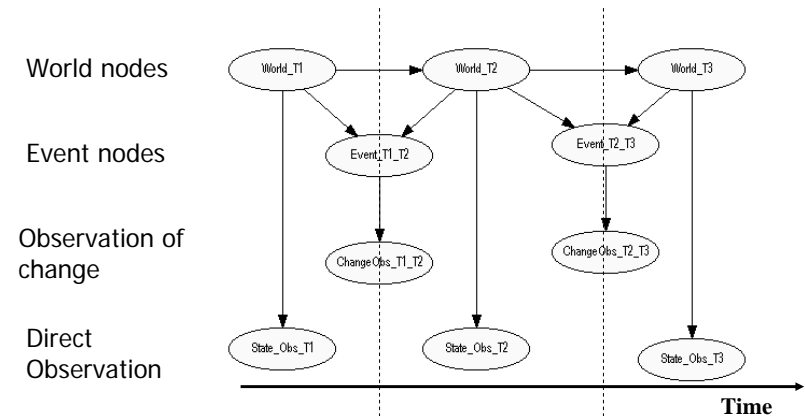
• **E** è l'insieme degli archi orientati:
 $E = \{(v, u) \mid v \in V(t_{i-1}), u \in V(t_i), \text{ con } i \leq n\}$

cioè è l'insieme degli **archi temporali** relativi alle fasce di tempo t_i , con $i = 1, \dots, n-1$

Gli archi temporali definiscono come le distribuzioni delle variabili al tempo t_i sono condizionalmente dipendenti dalla loro distribuzione al tempo t_{i-1} .

BN dinamici per il monitoraggio

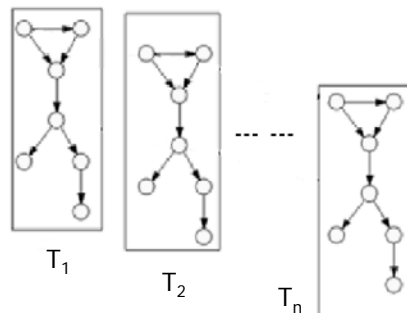
Il modello di Nicholson e Brady



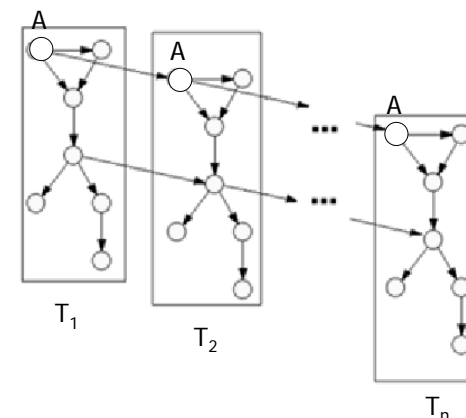
When the sensors indicate that something of interest in the World has changed, the network is extended with an additional time slice. ⁵⁸

Modelli dinamici 'senza effetto di trascinamento'

Tutte le variabili rappresentate nella rete hanno, al tempo T_i , una distribuzione di probabilità che non dipende dalla distribuzione al tempo T_{i-1}



Modelli dinamici 'con effetto di trascinamento'



La distribuzione di probabilità della variabile associata al nodo A dipende (oltreché dalle distribuzioni delle variabili associate agli altri nodi) anche dalla distribuzione di A al tempo precedente.

Si può simulare un *effetto decadimento*:

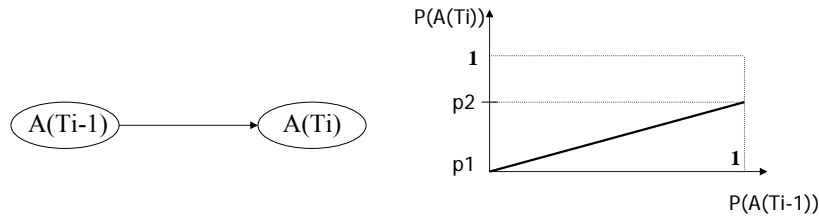
$\text{Prior}(A, T_i) < \text{Prior}(A, T_{i-1})$

Oppure un *effetto incremento*:

$\text{Prior}(A, T_i) > \text{Prior}(A, T_{i-1})$

Due possibili effetti di trascinamento: decadimento e incremento

Simulazione dell'effetto di decadimento



$$P(A(Ti) \mid \neg A(Ti-1)) = p1$$

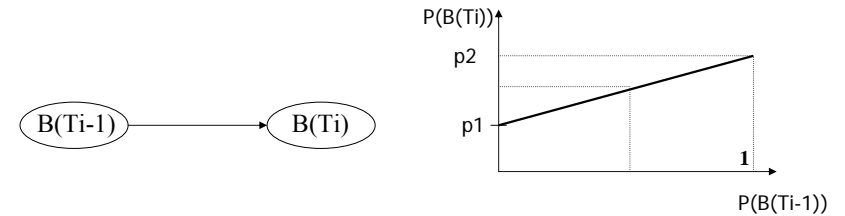
$$P(A(Ti) \mid A(Ti-1)) = p2$$

Variando i due valori di probabilità condizionata, si può ottenere un effetto di decadimento più o meno forte

Ad esempio: al nodo B posso associare una variabile che rappresenta lo "stato emotivo" di U (ipotizzando che le emozioni tendano a decadere nel tempo)

61

Simulazione dell'effetto di incremento



$$P(B(Ti) \mid \neg B(Ti-1)) = p1$$

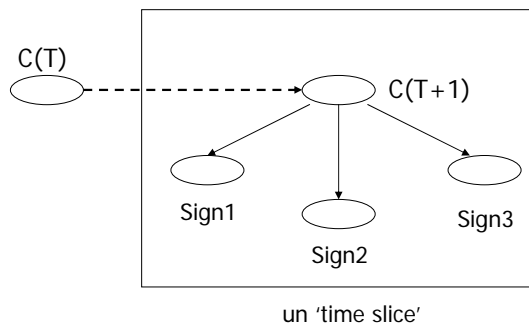
$$P(B(Ti) \mid B(Ti-1)) = p2$$

Variando i due valori di probabilità condizionata, si può ottenere un effetto di incremento più o meno forte

Ad esempio: al nodo B posso associare una variabile che rappresenta il 'grado di coinvolgimento' in un gioco da parte di U (ipotizzando che questo grado di coinvolgimento tenda a crescere man mano che il gioco procede nel tempo)

62

La variabile 'dinamica' può avere legami con altre variabili, in ogni singolo strato



C denota una caratteristica dell'utente da osservare 'nel tempo'.
 $C(T) \rightarrow C(T+1)$ è il legame 'fra strati diversi di tempo'

Sign1, Sign2 e Sign3 denotano i diversi modi in cui la variabile C si manifesta in un certo istante di tempo.

Ad esempio:
 C rappresenta lo 'stato di ansia'. Sign1 rappresenta il tono ansioso della voce, Sign2 il battito cardiaco accelerato, Sign3 i tratti 'tesi del volto, ...

63

Ragionare in un BN Dinamico per il monitoraggio

- Il tempo è discretizzato ad intervalli di ampiezza diversa, in corrispondenza dell'occorrenza di eventi esterni.
- Il BN al tempo t_i , $BN(t_i)$, rappresenta un modello del dominio al tempo considerato.
- Alcuni nodi N_j in $BN(t_i)$ dipendono non solo da altri nodi dello stesso network ma anche dal loro valore al tempo precedente $N_j(t_{i-1})$.
- I parametri associati al legame $N_j(t_{i-1}) \rightarrow N_j(t_i)$ misurano l'effetto di persistenza del valore di N_j da un istante di tempo al successivo.

64

Algoritmo per l'Espansione in un modello di monitoraggio (Nicholson&Brady)

1. Receive observation data indicating that change has occurred
2. Make new t_{i+1} instance of world and direct observation nodes
3. Connect old t_i and new t_{i+1} world and direct observation data
4. Create new event and change observation nodes
5. Connect world, event and change observation nodes
6. Add data as evidence for observation nodes
7. Run inference algorithm to update beliefs

65

Quali Tool possiamo utilizzare per la costruzione e la gestione di RCP?

- Hugin, per la costruzione manuale (impareremo ad usarlo nell'esercitazione di laboratorio)
- Bayesware – per l'apprendimento
- BNJ
- Netica
- MSBNX
<http://research.microsoft.com/adapt/MSBNx/>

... alcuni dei quali in versione free...

66

Riferimenti

Articoli reperibili sul sito:

- Spiegelhalter (per una introduzione alle RCP):
<http://www.di.uniba.it/intint/people/ArtDid/spiegelhalter.zip>
- Nicholson & Brady (per le RCP dinamiche):
<http://www.di.uniba.it/intint/people/ArtDid/nicholson.zip>
- Sebastiani (per l'apprendimento delle reti)
<http://www.di.uniba.it/intint/people/ArtDid/sebastiani.zip>
- Cooper & Herskovitz (algoritmo di apprendimento di una RCP)
<http://www.di.uniba.it/intint/people/ArtDid/bayesNet.zip>

67