

Laurea Specialistica in Informatica
a.a. 2006-2007

Interazione Uomo-Macchina II:

Interfacce Intelligenti

Valeria Carofiglio & Fiorella de Rosis

1

Introduzione

Prima parte: Formalizzazione e Ragionamento

1.1. Ragionamento logico:

- Formalizzazione
- Risoluzione

1.2. *Ragionamento incerto*

- ***Reti Causali Probabilistiche***
- Reti dinamiche
- Apprendimento di Reti

Programma
del Corso

Seconda parte: Modelli di Utente

2.1. Modelli logici

2.2. Modelli con incertezza

Terza parte: Interazione in linguaggio naturale

3.1. Generazione di messaggi

- Introduzione
- Teorie
- Metodi

3.2. Comprensione di messaggi

Quarta parte: Simulazione di dialoghi

Consideriamo di nuovo l'esempio Es 2.5

"Mangiare ad orari fissi **aiuta ad** evitare di saltare i pasti ..."
avavamo 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.
 $\forall x (\text{Person}(x) \wedge \text{EatAtFixedTime}(x)) \text{ ®? } \text{AvoidJumpMeal}(x)$

L'implicazione avrà, in questo modo, un *grado di forza*.₃

Consideriamo ora l'esempio Es 2.4

Abbiamo definito come 'certi' i dati:

Implies(a,g)

Likes(x,g)

CanDo(x,a)

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

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.

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

Cominciamo quindi con qualche chiarimento sul concetto di
probabilità.

Diversi Concetti di Probabilità

1. Definizione 'classica' (Laplace, 1820)

la probabilità, $p(E)$, di un evento E è il rapporto tra il numero di casi "favorevoli" al suo manifestarsi e il numero totale di risultati ugualmente possibili e mutuamente escludentesi:

$$p(E) = \frac{n^\circ \text{ casi favorevoli}}{n^\circ \text{ casi possibili}}$$

Viene anche detta probabilità a priori perché è calcolata a partire da "simmetrie" del problema, considerando equivalenti eventi "interscambiabili".

Esempio

In una moneta le due facce presenti, testa e croce, escono "indifferentemente"

, - $p(\text{"esce testa"}) = \text{numero di facce con testa} / \text{numero di facce totali}$

Diversi Concetti di Probabilità

1. Definizione 'classica' (Laplace, 1820)

Esempi

Estraiamo una carta da un mazzo da quaranta.

Calcoliamo la probabilità dei seguenti eventi:

$A = \text{"esce un sette"}$

$B = \text{"esce una figura"}$

$C = \text{"esce una carta di cuori"}$

$D = \text{"esce una figura di cuori"}$

$$P(A) = 4/40 = 1/10 = .1$$

$$P(B) = 12 / 40 = 3/10 = .3$$

$$P(C) = 10/40 = 1/4 = .25$$

$$P(D) = 3 / 40 = .075$$

$$0 \leq P(\text{Evento}) \leq 1$$

Diversi Concetti di Probabilità

1. Definizione 'classica' (Laplace, 1820)

Osserviamo che

$$0 \leq P(E) \leq 1$$

$$P(E) + P(\text{not } E) = 1$$

- se nessun caso possibile è favorevole al verificarsi dell'evento E (n° casi favorevoli = 0), E è detto **impossibile** e la sua prob. è 0.

Es: un'estrazione da un mazzo da quaranta carte,
sono impossibili gli eventi
“esce un jolly”, “esce il nove di cuori”;

- se tutti i casi possibili sono favorevoli al verificarsi dell'evento E (n° casi favorevoli = n), E è detto certo e la sua prob. è 1.

Es: un'estrazione da un mazzo da quaranta carte,
è certo l'evento
“esce una carta di cuori, o di denari, o di fiori, o di picche”;

7

Diversi Concetti di Probabilità

1. Definizione 'frequentista' (Von Mises, 1928)

Se, in una sequenza di n prove, un evento E si verifica s volte, si dice che il rapporto

$$freqRel(E) = \frac{s}{n}$$

è la **frequenza relativa** di E rispetto alla data sequenza di prove.

Esempio:

se estraendo 50 volte una carta da un mazzo da quaranta,
otteniamo 14 volte una carta di fiori,
diciamo che la frequenza relativa dell'evento
“si estrae una carta di fiori” è
 $14/50 = 0,28$.

8

Diversi Concetti di Probabilità

1. Definizione 'frequentista' (Von Mises, 1928)

la probabilità $p(E)$ di un evento è il limite cui tende la frequenza relativa del suo verificarsi all'aumentare del numero di esperimenti

$$p(E) = \lim_{n \rightarrow \infty} \frac{s}{n}$$

- è possibile effettuare esperimenti, sempre nelle stesse condizioni, e dove, almeno in linea teorica, il numero delle prove effettuate può essere reso grande a piacere;
- si dispone di numerosi dati statistici relativi ad un certo evento.

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

9

Diversi Concetti di Probabilità

1. Definizione 'frequentista' (Von Mises, 1928)

Esempio

Se lanciamo una puntina da disegno, essa può cadere con la punta rivolta verso l'alto o verso il basso.

Quali sono le probabilità dei due eventi?

È chiaro che sarebbe errato usare la definizione classica di probabilità:

a differenza della moneta,

la puntina **non è simmetrica**,

e uno dei due eventi può essere più probabile dell'altro.

10

Limiti delle concezioni 'classiche'

le due "definizioni" di probabilità appena incontrate hanno indotto a

1. ritenere che il **valore di probabilità sia oggettivo**, cioè che sia insito nella natura delle cose e non dipenda da chi lo valuta;
2. credere che si possa parlare di probabilità solo in questi due casi, molto particolari e riduttivi, rispetto alla complessità del mondo reale.

Sarebbero esclusi da argomentazioni probabilistiche
tutti quegli **eventi**

(la stragrande maggioranza di quelli di interesse pratico e scientifico)
per i quali è impossibile eseguire

l'inventario dei casi possibili e di quelli favorevoli

o per i quali non è possibile ripetere "infinite volte"
l'esperimento nelle identiche condizioni.

11

Limiti delle concezioni 'classiche'

Qual è la probabilità che questo pomeriggio piova?

Che domani io sia interrogato di italiano?

Che la Ferrari vinca la prossima gara di Formula uno?

Che il prezzo delle azioni Fiat salga nei prossimi due mesi di
almeno il 10%?

**Sembrerebbero non rientrare nel dominio dei problemi
trattati dalla teoria della probabilità.**

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

Eppure.....

12

Limiti delle concezioni 'classiche'

Qual è la probabilità che questo pomeriggio piova?
Che domani io sia interrogato di italiano?
Che la Ferrari vinca la prossima gara di Formula uno?
Che il prezzo delle azioni Fiat salga nei prossimi due mesi di
almeno il 10%?

**Eppure esiste una valutazione soggettiva della
probabilità:**

in base ad essa

prenderò oppure no l'ombrello,
studierò più italiano che matematica,
scommetterò o meno sulla vittoria della Ferrari,
acquisterò o venderò azioni Fiat.

13

Un'idea (relativamente) nuova di probabilità

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

***la probabilità $p(E)$ di un evento è grado di fiducia che una persona
coerente attribuisce al verificarsi di un evento***

Per rendere operativa questa definizione:

La probabilità di un evento E
è il prezzo che un individuo giudica equo pagare per ricevere un importo
unitario nel caso che l'evento E si verifichi.

Un individuo si dice **coerente**
se nessuna combinazione di scommesse eque a cui partecipa gli consente
di realizzare un guadagno o una perdita certa .

14

Un'idea (relativamente) nuova di probabilità

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

SOGGETTIVO NON VUOL DIRE ARBITRARIO

..ma

Legato alle conoscenze del soggetto

Valutazione tramite i bookmakers

Un cavallo dato "S a T" ha la probabilità

$$P = S/(S+T)$$

di vincere in un gioco equo.

La quota della posta "S a T"

**significa che si è disposti a ricevere T in caso di successo
e dare S in caso di insuccesso.**

15

Un'idea (relativamente) nuova di probabilità

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

La probabilità di un evento E

è il prezzo che un individuo giudica equo pagare per ricevere un importo unitario nel caso che l'evento E si verifichi.

Un individuo si dice coerente se nessuna combinazione di scommesse eque a cui partecipa gli consente di realizzare un guadagno o una perdita certa.

si basa su:

- ***una stima soggettiva di probabilità di un evento come 'somma che una persona è disponibile a scommettere sull'evento';***
- ***ipotesi di 'scambiabilità': che due scommettitori siano disponibili a scambiarsi i termini della scommessa;***
- ***ipotesi di 'non desiderio di perdere', che comporta la non disponibilità 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).¹⁶

Un'idea (relativamente) nuova di probabilità

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

Esempio

*Se Alice è disposta a scommettere 3 contro 4
sul fatto che si verifichi un certo evento,*

attribuisce a tale evento una probabilità pari a $3/(3+4)$ (circa il 43%).

Numeratore

quanto Alice è disposta a puntare

denominatore

somma della sua puntata

e di quella di un sfidante invocato a convalidare la valutazione.

*Tale somma rappresenta anche quanto ciascuno dei due participant i
alla scommessa vincerebbe a seguito della puntata.* 17

Bruno de Finetti: Definizione soggettiva e operativa di probabilità.

The classic exposition of his distinctive theory is the 1937 "La prévision: ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré*, 7, 1-68, which discussed *probability founded on the coherence of betting odds and the consequences of exchangeability.*

De Finetti ha proposto il seguente esperimento:

You must set the price of a promise to pay \$1 if there was life on Mars 1 billion years ago, and \$0 if there was not, and tomorrow the answer will be revealed.

You know that your opponent will be able to choose either to buy such a promise from you at the price you have set, or require you to buy such a promise from your opponent, still at the same price. In other words: you set the odds, but your opponent decides which side of the bet will be yours.

The price you set is the "operational subjective probability" that you assign to the proposition on which you are betting. This price has to obey the probability axioms if you are not to face certain loss, as you would if you set a price above \$1 (or a negative price). Prices, or equivalently odds, that do not expose you to certain loss are called *coherent*.

Vedi anche a

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

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

Vediamo un esempio

Qual è la probabilità che la ferrari vinca?

Io penso (credo, sono convinta...) che sia = .48

Questo significa che:

- 1) sono disponibile a pagare 48 euro, accettando di riceverne 100 (48 dati + 52 di vincita) se la ferrari perde e 0 se vince.
- 1) Sarei disponibile (per coerenza) a pagare 52 euro, accettando di riceverne 100, se la ferrari vince e 0 se perde.

Complessivamente, avremo raccolto 100 euro.

Quali sono i guadagni?

Se la ferrari vince, chi aveva pagato 52 euro ne guadagna 48;
se perde, chi aveva pagato 48 euro ne guadagna 52.

19

L'idea Soggettivista di probabilità rispetta le regole del Calcolo della Probabilità

Poiché l'ipotesi è che
la persona sia *disponibile a stare*
da ciascuna delle due parti che partecipano alla scommessa,
dovrà accettare di essere:

quello che paga la somma $P(E)$ e
che riceve 1 o 0 a seconda che E si verifichi o no, o viceversa.

Questo vincolo di coerenza implica che, se vi sono n eventi E_i esaustivi e mutuamente esclusivi ed una persona assegna una probabilità $P(E_i)$ ad ognuno di essi, debba essere d'accordo sullo scambiare

$\sum_{i=1, \dots, n} P(E_i)$ unità contro 1 unità.

Un insieme di valori di probabilità che non consente scommesse certamente perdenti si dice coerente.

20

Frequentismo e soggettivismo 'non radicali'

Un frequentista 'non radicale' si rende conto di dover prendere decisioni anche in condizioni di incertezza non misurabili attraverso frequenze.

In questi casi, piuttosto che inventare argomenti frequentisti difficili da sostenere, riconoscerà che le sue stime di probabilità corrispondono semplicemente a dei 'degree of belief' e cercherà di rispettare il vincolo di coerenza dei soggettivisti, per utilizzare le regole del calcolo delle probabilità.

D'altra parte, anche un 'soggettivista non radicale' è convinto che assegnare all'uscita (al lotto) del 53 una prob di 1/90 sia più corretto che assegnarle il valore di 1/80 (a meno che non abbia ragioni diverse per crederlo: " *Me lo ha detto mio nonno in sogno, e mio nonno mi ha sempre dato numeri buoni...*").

21

Inseriamo l'incertezza nella nostra conoscenza

Avremo un insieme di implicazioni in cui la relazione fra antecedente e conseguente è 'incerta':

$\forall x \forall a \forall g$

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

$\forall x \forall a \forall g$

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

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

Questo insieme di implicazioni è rappresentabile con un grafo, a ciascuno dei cui nodi è associato uno degli elementi dell'implicazione.

22

Vediamo allora come rappresentare questa forma di conoscenza e come ragionare secondo la teoria bayesiana della probabilità.

Considereremo, fra i vari metodi proposti, le *Reti Causali Probabilistiche*, che sono oggi il formalismo prevalente nel campo delle interfacce intelligenti.

(il loro uso è stato introdotto per la prima volta, in questo dominio, da un Centro Ricerche della Microsoft, con il '*Progetto Lumière*')

23

Definizione di Belief Network

Rappresentazione grafica delle **indipendenze condizionali**

Componenti:

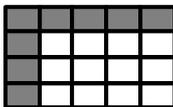


1. Insieme di variabili random: nodi discreti o variabili continue

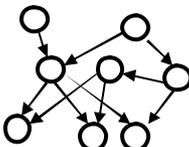


2. Insieme di archi/frecce orientati che connettono i nodi a rappresentazione delle dipendenze

Nota: l'assenza di un arco indica indipendenza



3. Una tavola di probabilità condizionata è associata ad ogni nodo (probabilità a priori)



4. Il grafo è a-ciclico (DAG: directed acyclic graph dalla teoria dei grafi)

24

Definizione di Belief Network

Se il grafo contiene le variabili $x_1, x_2, \dots, x_i, \dots, x_n$,
 e G_1 è l'insieme dei *genitori* di x_i ,
 una quantificazione completa e consistente dell'influenza
 fra le variabili può essere ottenuta specificando,

$$" x_i, \text{ la Prob}(x_i | G_i).$$

In questo caso, la *distribuzione congiunta di probabilità*
 $\text{Prob}(x_1, x_2, \dots, x_i, \dots, x_n)$ potrà essere approssimata con il
 prodotto delle probabilità condizionate:

$$\text{Prob}(x_1, x_2, \dots, x_i, \dots, x_n) = \prod_i \text{Prob}(x_i | G_i).$$

25

Indipendenza condizionale

Esempio1: "Le strade ghiacciate"

L'ispettore di polizia Smith e' impaziente per l'arrivo di Mr Holmes e del Dott. Watson. I due sono in ritardo l'ispettore Smith ha un altro appuntamento (pranzo)

Guardando fuori dalla finestra egli si domanda se le strade sono ghiacciate

Mr Smith sa che Mr Holmes e' Dr. Watson sono pessimi guidatori, cosi' se le strade sono ghiacciate, potrebbero fare un incidente.

La sua segretaria entra e gli dice che il Dr. Watson ha avuto un piccolo incidente d'auto.

"Watson?... Ok, sarebbe potuto andare peggio: le strade sono ghiacciate. Allora probabilmente anche Mr Holmes ha avuto un incidente. Ora andro' a mangiare!!"

"La strada ghiacciata?" replica la segretaria "Non e' cosi' freddo".

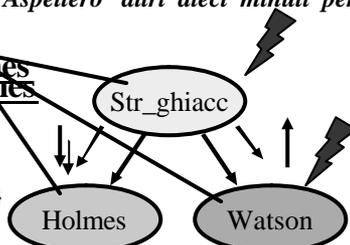
L'ispettore e' sollevato e dice:" Sfortuna per Watson. Aspettero' altri dieci minuti per Holmes!"

~~Modellizzazione dell'Indipendenza tra Watson e Holmes~~
 Modellizzazione dell'Indipendenza tra Watson e Holmes

Strada ghiacciata = "Str_ghiacc" → {vero, falso}

Holmes ha un incidente = "Holmes" → {vero, falso}

Watson ha un incidente = "Watson" → {vero, falso}



Due eventi dipendenti sono diventati indipendenti in seguito all'acquisizione di nuova informazione

Indipendenza condizionale

Esempio2: "L'erba bagnata"

'Mr Holmes ora vive a Los Angeles.

Un giorno uscendo da casa sua realizzo' che l'erba del suo giardino era bagnata. Questo era dovuto alla pioggia oppure aveva dimenticato lo spruzzatore aperto? Le sue credenze relative ad entrambi gli eventi aumentarono.

Successivamente egli noto' che l'erba del suo vicino, Dr. Watson, era anche bagnata.

Elementare: Holmes era quasi certo che avesse piovuto.'

Indipendenza tra Pioggia e Spruzzatore

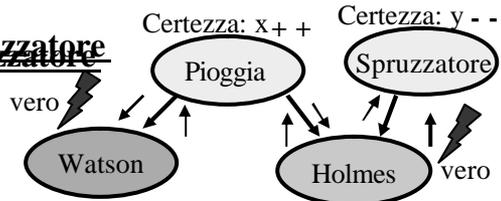
Modellizzazione :

Ha piovuto = "Pioggia" → {vero, falso}

Spruzzatore aperto = "Spruzzatore" → {vero, falso}

l'erba di Holmes è bagnata = "Holmes" → {vero, falso}

l'erba di Watson è bagnata = "Watson" → {vero, falso}



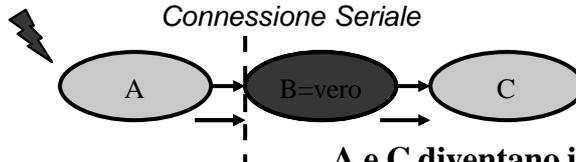
Questo processo viene chiamato spiegazione alla lontana (explaining away) .

Mostra dipendenze cangianti con l'acquisizione di informazione

Inferenze nei Belief Network

I tre modi in cui una evidenza può essere trasmessa in un BN

Connessione Seriale

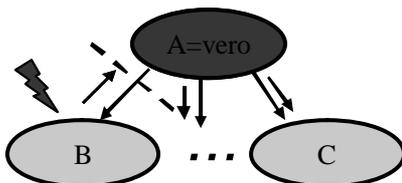


A e C diventano indipendenti

(d-separati dato B)

Evidenza diretta su A

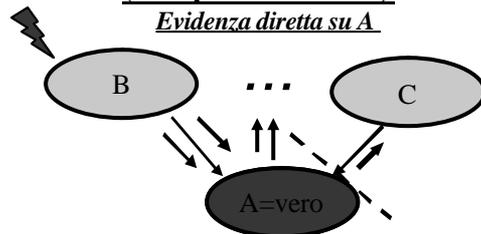
Connessione Divergente



B...C diventano indipendenti

(d-separati dato A)

Generalizzazione esempio 1



Se A è noto B..C..C
diventano indipendenti

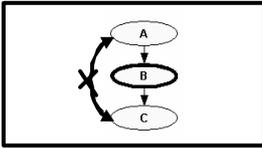
Generalizzazione esempio 2

D-separazione

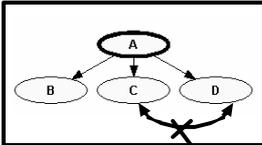
Sintesi

Se due variabili sono d-separate, sono condizionalmente indipendenti (nessuna evidenza passa)

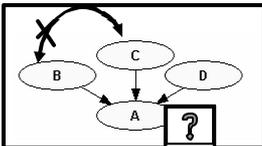
Due variabili sono d-separate se esiste una variabile intermedia tale che la connessione sia:



Seriale e lo stato della variabile intermedia è noto



Divergente e lo stato della variabile intermedia è noto



Convergente e lo stato della variabile intermedia è non noto

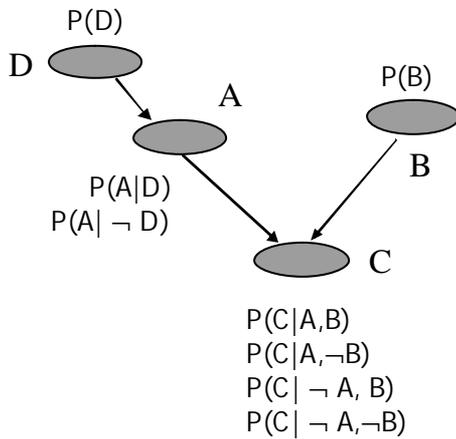
29

Fasi nella Costruzione di una RCP

1. Stabilire la differenza fra variabili 'nascoste' (hidden) e 'osservabili'
2. Definire la struttura del grafo:
 - semantica dei nodi
 - semantica delle relazioni
 - attenzione alle ipotesi di indipendenza condizionale!
3. Assegnare i parametri

30

Un esempio semplice di RCP

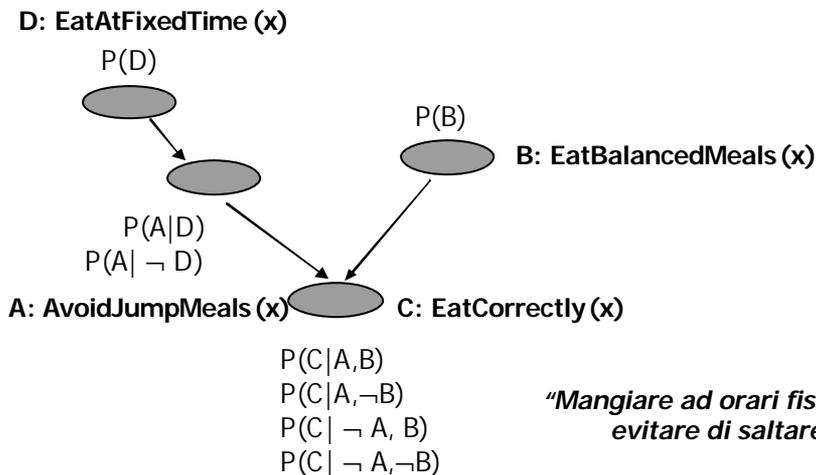


Il grafo rappresenta ipotesi di 'indipendenza condizionale' tra le variabili associate ai suoi nodi

ATENZIONE A CAPIRE
COME LE DIPENDENZE
CAMBIANO

31

Cosa potrebbe rappresentare la RCP che abbiamo appena visto



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

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

Quali parametri occorre specificare nella RCP

Le probabilità 'a priori' dei nodi-radice (quelli che non hanno genitori)

es: $P(D)$ (e $P(\neg D)$) $P(B)$ (e $P(\neg B)$)

Le probabilità condizionate di tutti i nodi intermedi e dei nodi - foglia (quelli che non hanno figli).

es: $P(C|A,B)$ (e $P(\neg C|A,B)$)

$P(C|A,\neg B)$

$P(C|\neg A, B)$

$P(C|\neg A,\neg B)$

33

Come assegnare i parametri alla RCP?

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

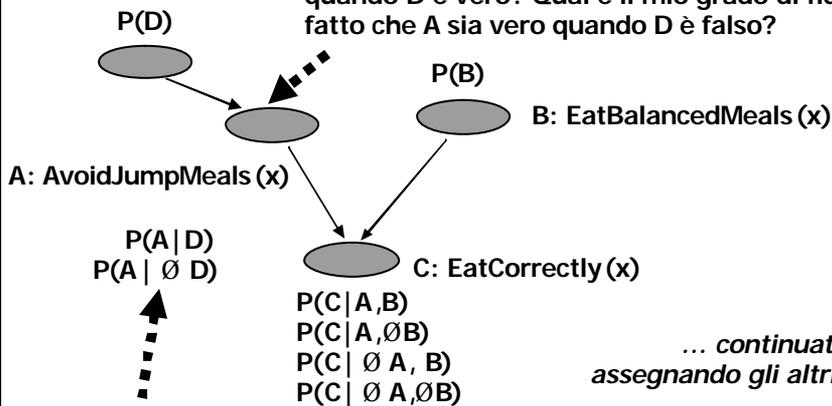
- a. Assegnare le prob a priori e condizionate sulla base della conoscenza soggettiva del dominio, e quindi del proprio '*grado di fiducia*' sull'associazione fra le variabili;
- b. 'Calcolare' questi parametri da una *misura di frequenze* su un set di dati, applicando metodi di apprendimento.

Come vedremo, i due metodi possono essere integrati con una procedura di *apprendimento interattivo della rete*.

34

Vediamo allora, ad esempio, come assegnare parametri alla nostra rete (Metodo a)

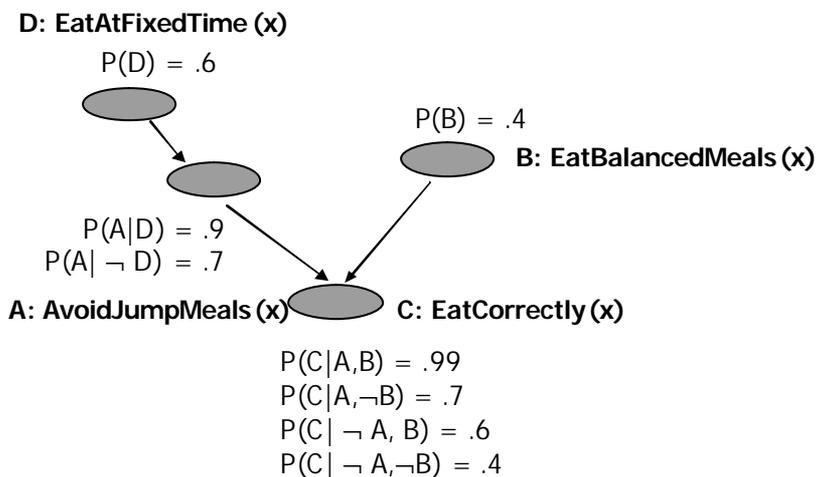
D: EatAtFixedTime (x) Qual è il mio grado di fiducia sul fatto che A sia vero quando D è vero? Qual è il mio grado di fiducia sul fatto che A sia vero quando D è falso?



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

35

... Continuiamo l'esempio



Le inferenze nella RCP

Assegnare un valore a
nodi-radice (probabilità incondizionate)
nodi intermedi e dei nodi -foglia (probabilità condizionate).

significa, implicitamente,

*assegnare valori di **probabilità a priori***
a ciascuna delle variabili.

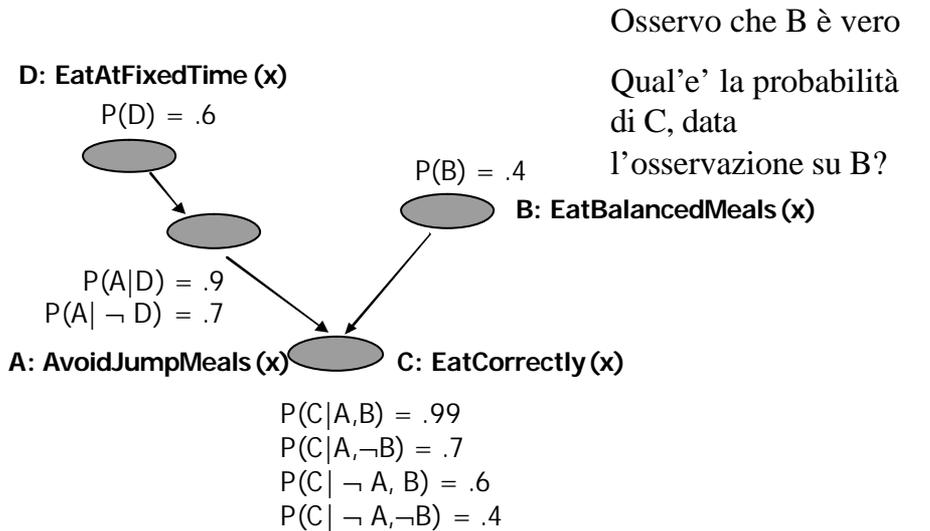
37

Le inferenze nella RCP

- 1) *"osservare" un fenomeno*
- 2) *Assegnare un valore (evidenza certa/incerta)*
a variabili della rete
- 3) *Applicare un meccanismo di propagazione delle evidenze*
*Per il calcolo dei valori di **probabilità a posteriori** $P^*(\cdot)$ di*
ciascuna delle variabili.

38

... Continuiamo l'esempio



ATTENZIONE ALLA DIFFERENZA TRA **PROBABILITA' A POSTERIORI** E **PROBABILITA' CONDIZIONATE** ³⁹

Verifiche sulle RCP

- Calibrare i parametri:
 - calcolare la prob a priori dei nodi intermedi e dei nodi-foglia
 - valutarne la plausibilità
 - testare la sensibilità del modello a variazioni più o meno piccole nei parametri e
 - rivedere i parametri critici
- Verificare la struttura:
 - inserire/cancellare archi rivedendo le ipotesi di indipendenza condizionale

Vediamo un algoritmo per la manipolazione dei parametri nelle RCP

(D. Spiegelhalter)

Propagazione dell'Evidenza in un BN

Il problema della propagazione corretta dell'evidenza in una RCP è NP-hard *

Un buon algoritmo di propagazione approssimata dell'evidenza (di Lauritzen e Spiegelhalter) opera sull'albero delle cricche (*junction tree*) anziché sul grafo iniziale.

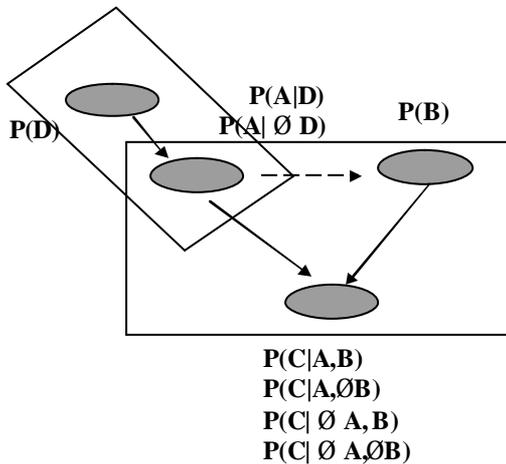
* G Cooper: "The computational complexity of probabilistic inference using bayesian belief networks", *Artificial Intelligence*, 1990

Vediamo come
un tool per la gestione di BN
calcola le 'probabilità a priori' dei nodi

e come
'propaga' nuova evidenza nella rete.

Vedremo poi quale algoritmo viene
applicato per realizzare questi compiti.

Outline dell'Algoritmo di Spiegelhalter



Il calcolo delle probabilità a priori 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 di ogni nodo** in funzione delle prob delle cricche e delle loro intersezioni.

Le **prob marginali del grafo** in funzione del rapporto fra:

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

Vediamo meglio:

43

Introduciamo il concetto di *probabilità marginale*

La **probabilità marginale** di un grafo è la **probabilità congiunta** dei valori assunti dai suoi nodi.

Nel nostro esempio: $P(A,B,C,D)$, oppure $P(\neg A,B,C,D)$, $P(A,\neg B,C,D)$, $P(\neg A,\neg B,C,D)$, ... ecc

Sappiamo che

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

Ma, per le ipotesi di indipendenza condizionale introdotte implicitamente nel grafo, possiamo semplificare:

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

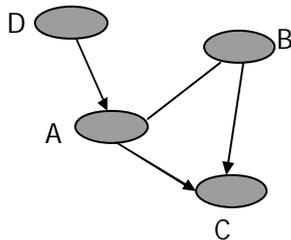
44

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:



In questo nuovo grafo, ABC e DA sono cricche.
A è l'intersezione fra di esse.

45

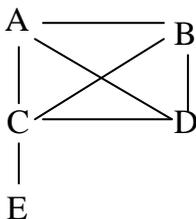
Cosa è una 'cricca'

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 distribuzione congiunta di probabilità su G può essere espressa in funzione delle distribuzioni congiunte delle sue cricche, in un particolare grafo non orientato

Probabilità a priori dei nodi

Prima abbiamo visto come calcolarle in funzione dei parametri inseriti nella rete:

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

Ma (A, D) è una cricca, così come (A, B, C) .

Quindi, in generale, le probabilità a priori dei nodi si possono calcolare a partire dalle probabilità delle cricche:

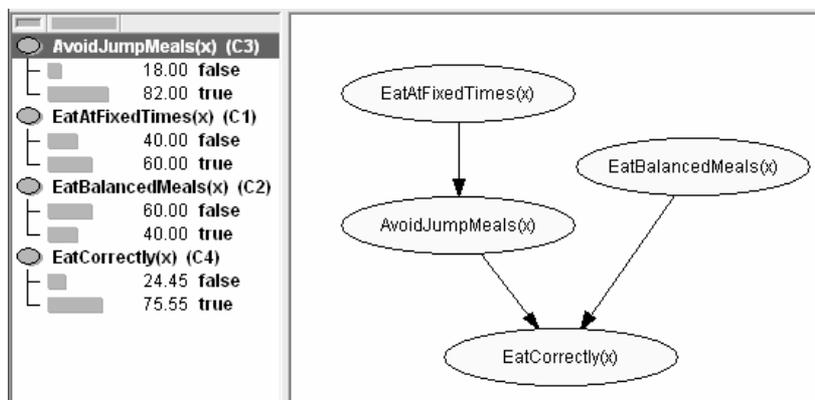
$$P(A) = P(A, B, C) + P(A, \neg B, C) + P(A, B, \neg C) + P(A, \neg B, \neg C)$$

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

... ecc

47

Probabilità a priori dei nodi nel nostro esempio



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

48

Tracing dell'algoritmo in Hugin

Il file .hlg

Marriages (bigamy allowed):
Marrying C2 and C3

Triangulation by minimum fill-in weight heuristic:
Fill-in links and node numbering:

4 C4
3 C2
2 C3
1 C1

Nota:
C1 è D
C2 è B
C3 è A
C4 è C

Cliques:

Clique 1, 2 members (C1, C3), table size = 4 ← (A,D)
Clique 2, 3 members (C3, C2, C4), table size = 8 ← (A,B,C)

Total clique table size: 12

The junction forest:

Creating junction tree with clique 1 as root ...

Cliques 2 and 1 linked, separated by {C3} (table size = 2)

Checking tables for all nodes ...

↑

49

A

Assignment of potentials to cliques:

Node C4 assigned to clique 2

Node C3 assigned to clique 1

Node C2 assigned to clique 2

Node C1 assigned to clique 1

Checking tables for all nodes ...

Mon Jan 30 13:53:09 2006

Beliefs for chance nodes:

C4 --

false 0.244480

true 0.755520

C3 --

false 0.180000

true 0.820000

C2 --

false 0.600000

true 0.400000

C1 --

false 0.400000

true 0.600000

50

Calcolo delle 'probabilità marginali'

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

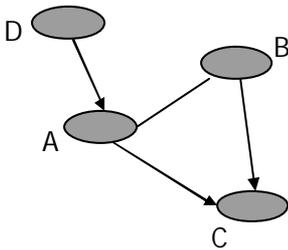
Semplificando:

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

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

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

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



Distribuzione congiunta di probabilità =
rapporto fra:
prodotto delle probabilità delle cricche
e
prodotto delle probabilità delle loro intersezioni

51

Vediamo ora come utilizzare una RCP

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

52

L'algoritmo di Spiegelhalter per i BN

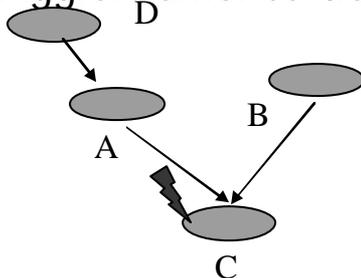
Il processo di propagazione dell'evidenza nella junction tree (invece che nel grafo di partenza)

- si basa sul ***passaggio di messaggi*** tra cricche adiacenti.
- L'approccio standard consiste nello scegliere una cricca come radice della junction-tree e, di seguito, procedere attraverso due fasi distinte:
 - ***Prima fase***: l'evidenza viene propagata tra i nodi della cricca-radice.
 - ***Seconda fase***: l'evidenza viene distribuita dalla cricca-radice alle cricche adiacenti secondo un ordine prestabilito:
 - ***riorientamento*** del grafo precedentemente moralizzato
 - ***un ordinamento delle cricche***. (criterio di massima cardinalità)

53

Prima fase:

Aggiornamento della probabilità delle cricche



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

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.

Consideriamo la cricca AD.

$$\begin{aligned} P^*(A,D) &= P(A,D|C) = P(A,D,C) / P(C) = \\ &= P(D|A,C) * P(A|C) = \\ &= P(D|A) * P(A|C) = \\ &= P(A,D) / P(A) * P(A|C) \\ &= P(A,D) * (P^*(A) / P(A)) \end{aligned}$$

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.

54

Probabilità a posteriori dei nodi

Si può calcolare dalla probabilità a posteriori delle cricche:

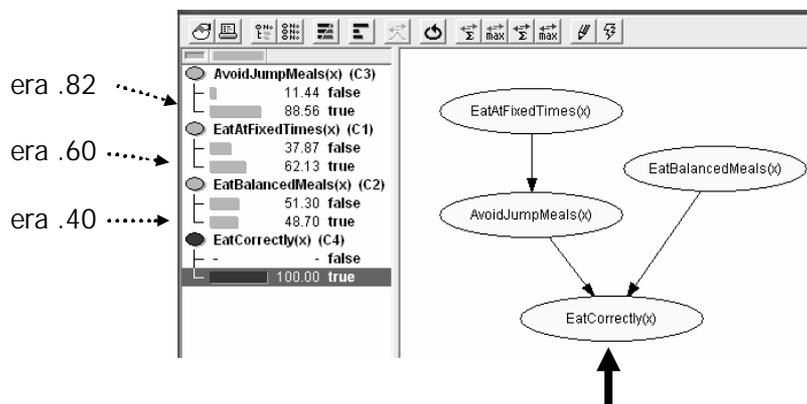
$$P^*(A) = P^*(A,B,C) + P^*(A,\neg B,C) + P^*(A,B,\neg C) + P^*(A,\neg B,\neg C)$$

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

... ecc

55

Nel nostro esempio



So che l'individuo considerato '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?

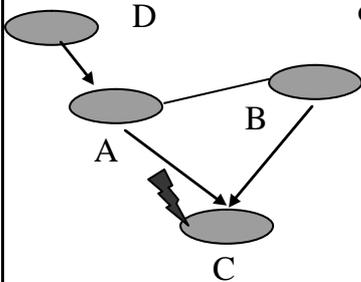
56

Criterio di massima cardinalità

Criterio di Massima cardinalità:

- definisce un metodo per **etichettare un grafo**.
- In questo contesto, fornisce un metodo per **ordinare i nodi nel grafo moralizzato**:

- assegna etichette ordinatamente a partire dal nodo su cui è inserita l'evidenza, e procedendo etichettando ad ogni passo quel nodo connesso al maggior numero di nodi già etichettati.

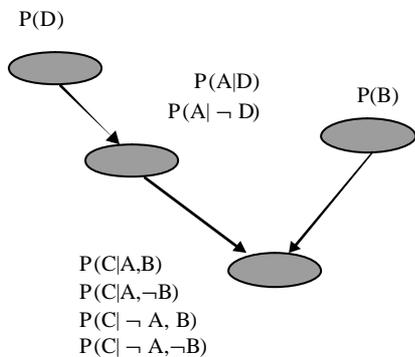


viene stabilito il seguente ordine sui nodi del grafo:

C,B,A,D \longrightarrow [ABC],[AD] (corrispondente ordine sulle cricche)

57

Forme di ragionamento

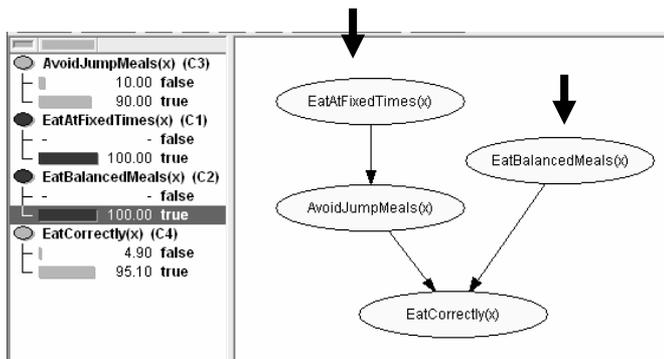


Si possono valutare le conseguenze, sull'intero grafo, di 'evidenza' acquisita su uno o più nodi:

- 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)

58

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??

59

Riprendiamo ora l'esempio della persuasione (Es 2.4)

- **Task model: le strategie persuasive di Walton**
 - a. *Appeal to positive consequences*
"Se ritieni che compiere una determinata azione comporti conseguenze importanti per te e sei in grado di compierla, dovresti farla".
 - b. *Appeal to negative consequences*
"Se ritieni che compiere una determinata azione comporti conseguenze che desideri evitare e puoi evitarle di compierla, dovresti farlo".
- **Domain model**
"Fare sport fa bene alla salute e alla forma fisica. Fare una vita sedentaria aumenta il rischio di ingrassare e di perdere tono muscolare".
"Il running è un particolare tipo di sport". Tutti i giovani senza particolari problemi di salute sono in grado di fare running.
- **User model**
Young(G)
Healthy(G)
Likes(G, InShape)

60

Come avevamo formalizzato il problema

- **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**

$\text{Young}(G)$

$\text{Healthy}(G)$

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

.....

61

Inseriamo l'incertezza

- Task model: le strategie persuasive di Walton

a. *Appeal to positive consequences*

"Se sei abbastanza convinto che compiere una determinata azione comporti conseguenze importanti per te e *pensi che potresti* compierla, allora *probabilmente* dovresti farla".

b. *Appeal to negative consequences*

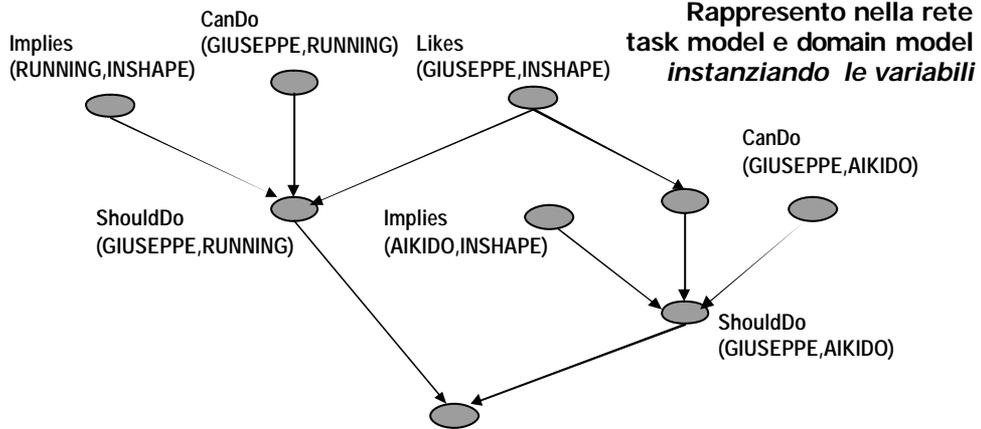
"Se sei abbastanza convinto che compiere una determinata azione comporti conseguenze che desideri evitare e *pensi che potresti* evitare di compierla, allora *probabilmente* dovresti farlo".

- Domain model

"Fare sport fa, *in genere*, bene alla salute e alla forma fisica. Fare una vita sedentaria aumenta, *in genere*, il rischio di ingrassare e di perdere tono muscolare".

"Il running è un particolare tipo di sport". *La maggior parte dei giovani senza particolari problemi di salute sono in grado di fare running.* ⁶²

Come formalizzare il problema con una RCP



Ho perso in generalità ma...

ShouldDo (GIUSEPPE,s)

...posso scegliere l'argomento più forte, data la conoscenza sul dominio e su Giuseppe, con il grado di incertezza di ogni elemento!!

63

Come cambia l'algoritmo di ricerca della strategia 'piu' promettente?

- Introduco nella rete l'evidenza (conoscenze che ho sul dominio e su Giuseppe)
- Propago
- Osservo i genitori del nodo 'ShouldDo(Giuseppe,s)'
- Cerco quello con probabilità massima: questo corrisponde alla strategia 'ottimale'.

Che messaggi posso generare?

- *Quale sport* deve fare Giuseppe
- *Perché* : osservando i genitori di questo nodo

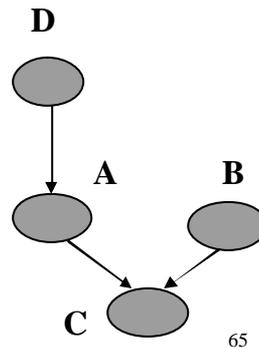
64

Traduzione dei risultati in LN

Posso tradurre il risultato della esplorazione della RCP in un messaggio in linguaggio naturale, utilizzando le informazioni associate ai nodi:

- nome
- tipologia
- testo associato
- tabelle di probabilità condizionata
- modalità di acquisizione dell'evidenza, ecc

"C is (likely/most likely,...) true
because A is very likely, as
You told me that D is true"



65

Esercizio

Completate l'esempio della persuasione, estendendo la struttura della rete, inserendo i parametri nel modello, simulando diverse situazioni possibili e ragionando sui risultati

66

Riferimenti

- Introduzione ai BN: articolo di Spiegelhalter reperibile sul sito
- Descrizione approfondita dei metodi: libro di Judea Pearl
- Argomenti specifici:
 - Il Progetto Lumiere della Microsoft:
research.microsoft.com/~horvitz/lumiere.HTM
 - applicazione alla gestione di modelli di Studente:
articoli di Kurt VanLehn
 - applicazione alla costruzione di modelli dell' Interlocutore nel dialogo: Ingrid Zukerman (Università di Monash, Australia)
 - applicazione alla traduzione di stati mentali in 'espressioni':
Gene Ball (Microsoft)
 - applicazione alla modellizzazione cognitiva di stati emotivi: articoli nostri reperibili sul sito