

Laurea Specialistica in Informatica
a.a. 2004-2005

Interfacce Intelligenti

Fiorella de Rosis

Programma del Corso

1. Introduzione
2. Ragionamento logico: formalizzazione
3. Il principio di risoluzione: usi e strategie diversi
4. *Ragionamento incerto: reti causali probabilistiche*
5. Modelli di utente
6. Interazione in linguaggio naturale: generazione di messaggi
7. Ipermedia adattivi
8. Interazione in linguaggio naturale: simulazione di dialoghi
9. Interazione con Agenti Animati
10. Prospettive: affective computing

Costruzione di Reti Causali Probabilistiche

Passi necessari:

- a. Definizione della struttura
- b. Assegnazione dei parametri
- c. Test
- d. Eventuale revisione iterativa del modello

Costruzione di Reti Causali Probabilistiche

Metodi possibili:

- a. Costruzione 'manuale' del modello (struttura e parametri), a partire da ipotesi sul dominio
- b. Apprendimento (dei parametri e/o della struttura) da un insieme di dati.
- c. Combinazione delle due procedure

Nel primo caso, la definizione dei parametri è basata sull'idea soggettivista di probabilità (degree of belief).

Nel secondo caso, è basata sulla sull'idea frequentista.

Nel terzo caso, su un'integrazione dei due concetti.

In ogni caso, il test del modello e la sua revisione iterativa sono passi necessari!

Apprendimento da un insieme di Dati: Passi necessari

- a. 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)
- b. Codifica dei dati (foglio Excel, file di testo,...) e verifica degli errori
- c. Discretizzazione delle variabili continue
- d. Calcolo delle prob a priori e condizionate
- e. Valutazione dell'*attendibilità* del modello
- f. Test preliminare del modello
- g. Eventuale revisione iterativa della struttura e dei parametri
- h. Test finale del modello in situazioni 'significative'

Apprendimento di RCP

Un Esempio:

Un modello degli atteggiamenti rispetto allo sport
(da utilizzare per la generazione di messaggi persuasivi)

Popolazione coinvolta nell'esperimento: gli studenti del Corso di IntInt

(circa 100 studenti, omogenei per livello culturale, abbastanza equidistribuiti per sesso ed età, diversi per provenienza: Bari-non Bari, personalità, ecc;

Variabili da osservare:

- variabili che definiscono il 'soggetto': età, sesso, luogo di residenza,
- interesse nei confronti dello sport
- sport praticati

Il questionario

Quanti anni hai? __

Qual è il tuo sesso? M __ F __

Dove risiedi? A Bari __

A meno di 100 km da Bari __

A più di 100 km da Bari __

Credi che essere in buona forma fisica sia importante?

molto __ abbastanza __ poco __

Credi che la salute sia un bene fondamentale della vita?

molto __ abbastanza __ poco __

Pratichi qualche sport?

con regolarità __ raramente __ mai __

Se sì, quale?

corsa__ bicicletta __ palestra __ nuoto__

Se raramente o mai, perché?

non m'interessa __ non ho tempo __ non ho soldi __

.....

Somministrazione ad un campione rappresentativo

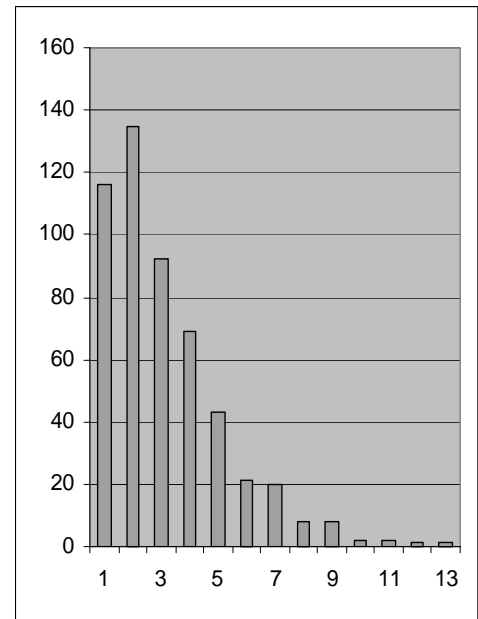
Trascrizione dei risultati in un database

Età	Sesso	Resid	Forma fisica	Salute	Pratichi sport	Quale	Perché no	...
22	M	B	Y	N	Y	PI	T	...
25	F	P	N	N	Y	PA	S	...

Discretizzazione delle variabili continue (es età):

- a. per intervalli di eguale ampiezza ('range');

es: in una variabile con distribuzione asimmetrica dei valori nell'intervallo (0, 200), una discretizzazione per 'range' produrrebbe alcuni 'dati sparsi', cioè alcuni valori con bassa probabilità



Discretizzazione delle variabili continue:

- b. per intervalli con eguale numero di casi ('frequency').

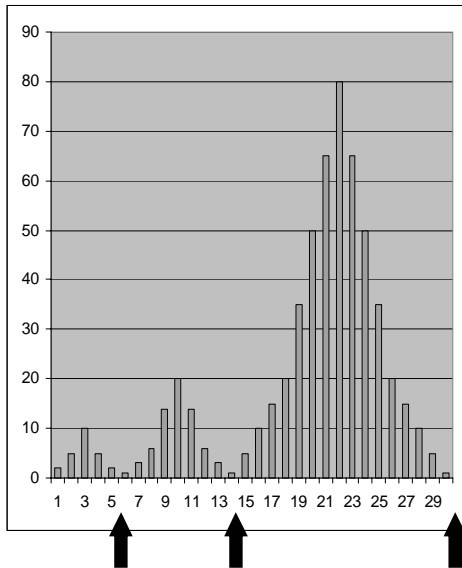
Nello stesso esempio precedente, una distribuzione per frequenza, con lo stesso numero di valori (13) produrrebbe una distribuzione 'uniforme' con i seguenti intervalli (di ampiezza crescente):

2-6, 6-10, 10-16, 16-21, 21-25, 25-30, 30-36, 36-43, 43-51, 51-59, 59-72, 72-100, 100-319

I due metodi presentano vantaggi diversi

Discretizzazione delle variabili continue:

Consideriamo la distribuzione seguente:



che potrebbe rappresentare, ad esempio, la distribuzione per età di una popolazione con tre classi di soggetti.

Qui, può essere conveniente una discretizzazione in tre valori che rispetti la distribuzione in classi

Calcolo delle Probabilità (a priori e condizionate) con il metodo 'frequentista'

Consideriamo il caso di due variabili, V_h , V_k

la prima a due e la seconda a tre valori:

	V_k^1	V_k^2	V_k^3
V_h^1	a	b	c
V_h^2	d	e	f

$$P(V_h^1|V_k^1) = a / (a+d)$$

... ecc

$$P(V_h^1) = (a+b+c)$$

... ecc

Ambiguità del valore .5 di probabilità

Una prob = .5 può significare 'incertezza' (su valori osservati) o 'ignoranza' (su valori non osservati)

Diagram illustrating the ambiguity of the value .5 in probability. Two windows, 'F-Farw', are shown. The left window displays 'Prob condizionate' (Conditional Probabilities) and the right window displays 'Freq Osservate' (Observed Frequencies). Both windows have columns: Qagt, Mtyp, Y, and N. The left window shows conditional probabilities, with several cells in the Y column circled (0.500, 0.500, 0.000, 0.500, 0.001, 0.000, 0.000, 0.002, 0.000, 0.009, 0.002, 0.958, 0.000, 0.019). The right window shows observed frequencies, with several cells in the Y column circled (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 23.0, 0.0, 4.0). A central oval labeled 'F-Farw' has arrows pointing to both windows.

Prob condizionate

Freq Osservate

Calcolo delle Probabilità (a priori e condizionate):

Notare che:
quanto maggiore è il numero di variabili messe in relazione fra loro (e/o il numero di valori che ciascuna di esse può assumere), tanto più complessa è la tabella delle probabilità condizionate.

Qagt

Back	Frship	Y	N
H	H	0.115	0.885
H	M	0.000	1.000
H	L	0.500	0.500
C	H	0.043	0.957
C	M	0.040	0.960
C	L	0.000	1.000

Back: 2 valori; Frship: 3 valori

F-Farw

Qagt	Mtyp	Y	N
Y	Su	0.500	0.500
Y	Sp	0.500	0.500
Y	Co	0.000	1.000
Y	Cu	0.500	0.500
Y	Fa	0.001	0.999
Y	An	0.000	1.000
Y	Qu	0.000	1.000
N	Su	0.002	0.998
N	Sp	0.000	1.000
N	Co	0.009	0.991
N	Cu	0.002	0.998
N	Fa	0.958	0.042
N	An	0.000	1.000
N	Qu	0.019	0.981

Qagt: 2 valori; Mtyp: 7 valori

Notare i valori 'sparsi' e i casi con P = .5!!

Apprendimento di RCP

Il problema dell'apprendimento del BN:

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).

Per semplificare tutti gli algoritmi di gestione del BN, è necessario determinare *il BN con il numero minimo di archi che rappresenta in modo 'accettabile' il dominio considerato.*

Apprendimento di RCP

Il problema dell'apprendimento:

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).

Apprendimento di RCP

Una misura di qualità del modello M:

La *probabilità marginale* (PM) di una RCP è il prodotto delle probabilità marginali dei suoi nodi.

Marginal Log Likelihood di un BN:

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

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.

Apprendimento di RCP

Ipotesi di approssimazione:

Le osservazioni sono distribuite in modo indipendente:

$$P(D | M) =$$

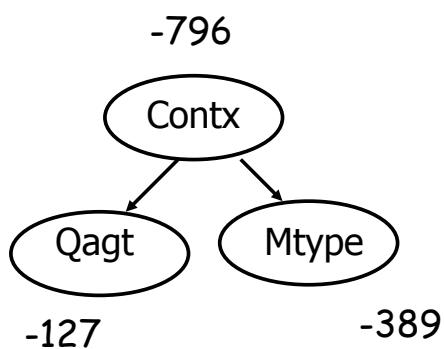
$$\prod_k \prod_h P(V_h = v_h^k | \text{Genitori}(V_h, M))$$

... i prodotti diventano somme se si passa ai logaritmi; questo rende le elaborazioni più semplici e aiuta a superare il problema dei numeri molto piccoli.

$$MLL(M) = \sum_k \sum_h \log P(V_h = v_h^k | \text{Genitori}(V_h, M))$$

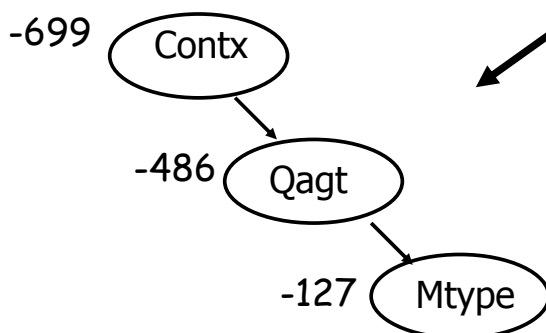
La MLL di una RCP è quindi uguale alla somma delle MLL dei suoi nodi.

Dipende dal numero di nodi nella RCP e (in misura molto minore) dalla sua struttura.



Un esempio

$MLL(RCP) = -1313$
in entrambi i casi



La MLL è una misura importante per *confrontare modelli alternativi* sullo stesso set di dati, più che per valutare singoli modelli

Apprendimento di RCP

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.

Apprendimento di RCP

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').

Apprendimento di RCP

La MLL di una RCP è funzione della complessità della rete (numero di nodi e di archi).

Il confronto fra due RCP apprese deve quindi tener conto:

- della loro complessità e, a parità di complessità,
- della loro plausibilità teorica e
- dei vantaggi ai fini dell'uso previsto.

Apprendimento di RCP

Riprendiamo l' esempio iniziale

Un modello delle preferenze negli spettacoli televisivi.

Età	Sesso	Resid	Forma fisica	Salute	Pratichi sport	Quale	Perché no	...
22	M	B	Y	N	Y	PI	T	...
25	F	P	N	N	Y	PA	S	...

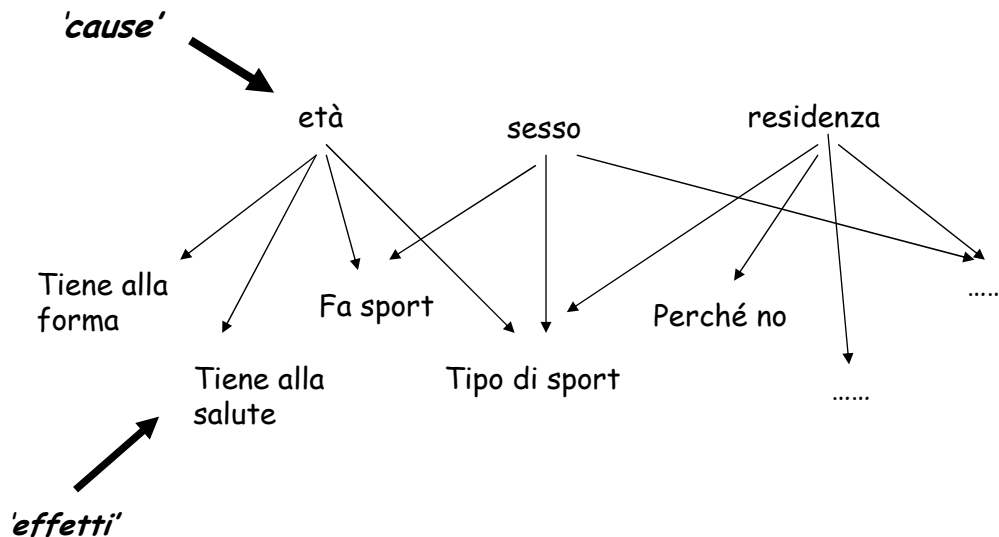
Numero possibile di nodi nel BN: 8

N massimo possibile di archi: 28

Ma esiste una struttura 'teorica',
che tenga conto delle ipotesi sul dominio?

Apprendimento di RCP

Possiamo distinguere le variabili in due grandi categorie,
Definendo la struttura del BN ...



Apprendimento di RCP

Possiamo affidare l'intero compito di
definizione della struttura e apprendimento dei
parametri all'algoritmo

Oppure

Possiamo adottare una strategia 'mista':

Definizione di 'vincoli' iniziali,

Completamento della struttura con algoritmo di
apprendimento,

Revisione iterativa della struttura e dei parametri.

Vediamo come....

Apprendimento di RCP

1 Caso: apprendimento (*quasi*) completamente automatico

Si può isolare un *sottoinsieme delle variabili*, su cui effettuare l'analisi (ad esempio, eliminando l'identificatore del soggetto).

Si possono *imporre vincoli* su:

- Il *numero di classi* in cui discretizzare le variabili continue e il *criterio* da applicare (per range o per frequenza).
- L'*ordine* di considerazione delle variabili: questo determina un ordine di considerazione dei potenziali genitori per ogni nodo.
- Il *numero massimo di 'genitori'* per ogni nodo.

Un esempio più complesso

Dialoghi con un Agente Conversazionale.

Variabili implicate:

Relative all'Utente: Gender, Background

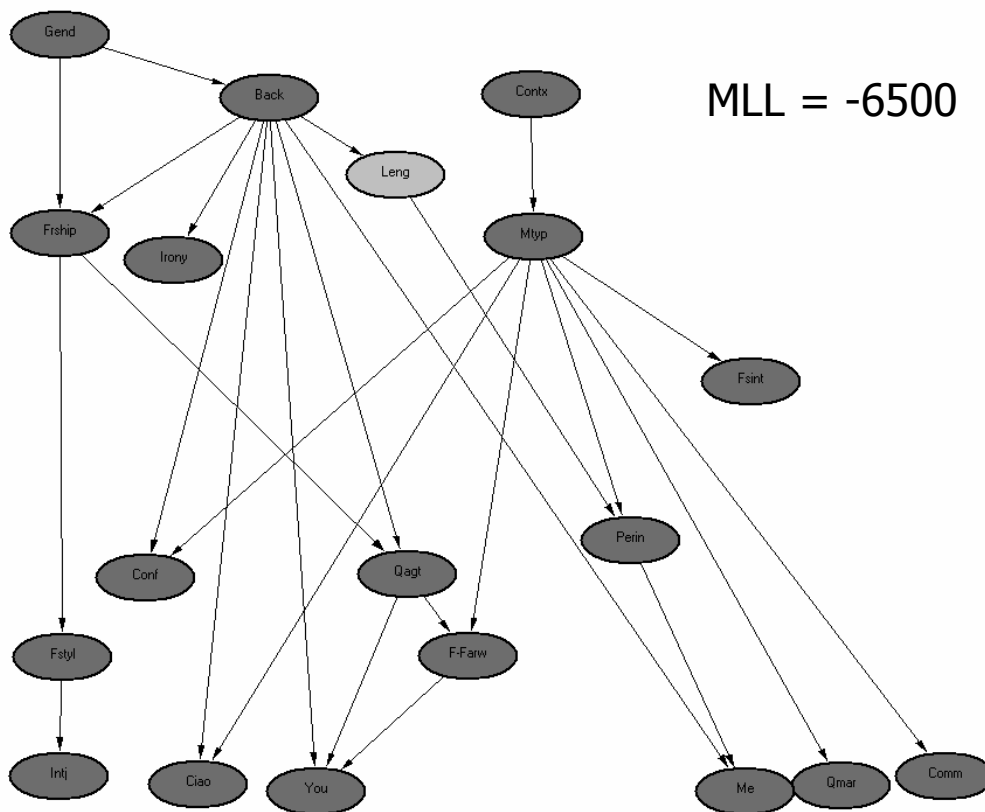
Relative al dialogo nel suo complesso: Friendship

Relative alla singola mossa di dialogo: Contx, Mtype, Length, PersInf, Friendly-SelfIntr, Fam-Style, Q-Agent, Irony, Friendly-Farewell

Relative alle caratteristiche linguistiche della mossa: Me, You, QuestMark, Interj, Ciao, FavComment, Confirm

Database: 30 soggetti. 705 'osservazioni' (mosse di dialogo)

La Rete Appresa



Vincoli:
Sull'ordine:
Var. di utente
Var. di dialogo
Var di mossa
Var linguistiche

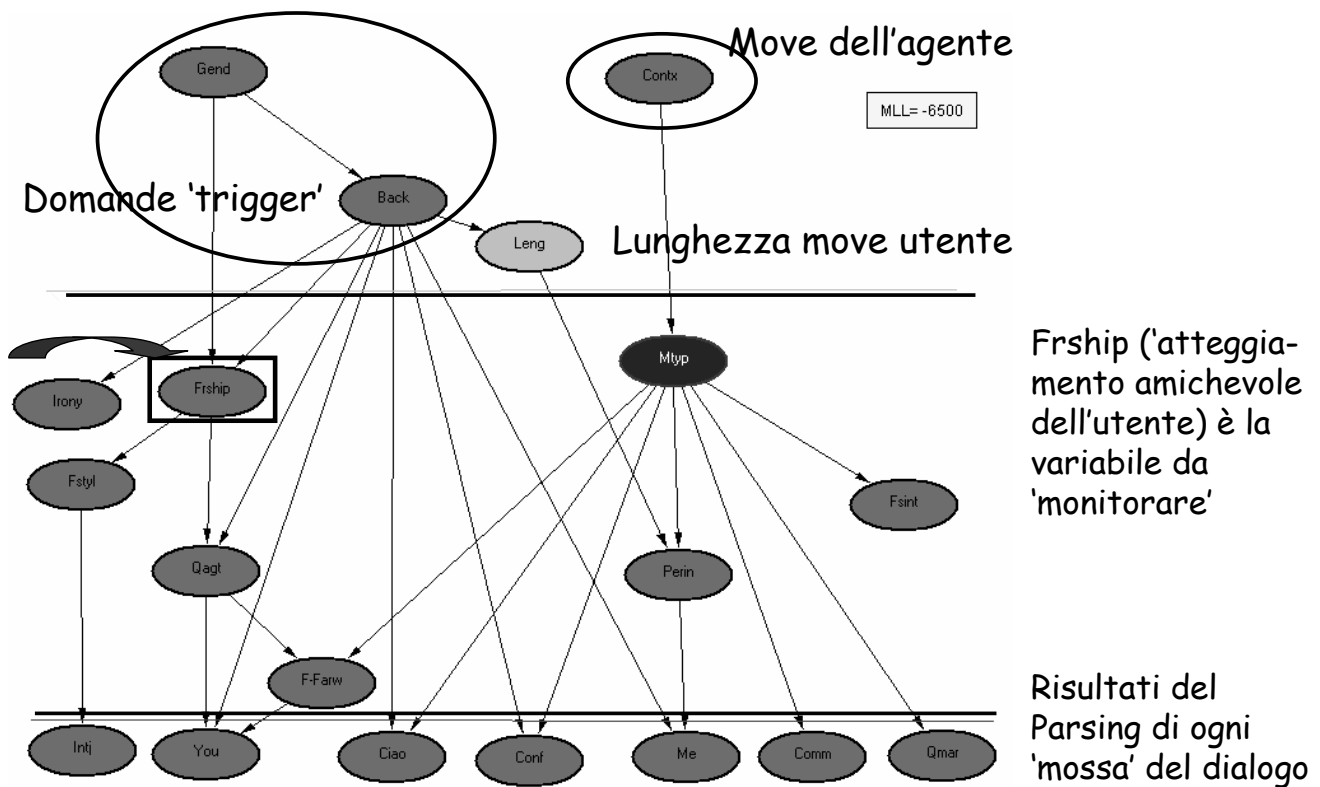
Discretizzazione
di Length per
frequenza

Nessun limite
Sul numero di
genitori

Come verificare
se la rete 'appresa' è soddisfacente????

Apprendimento di RCP

Come penso di usare la rete?

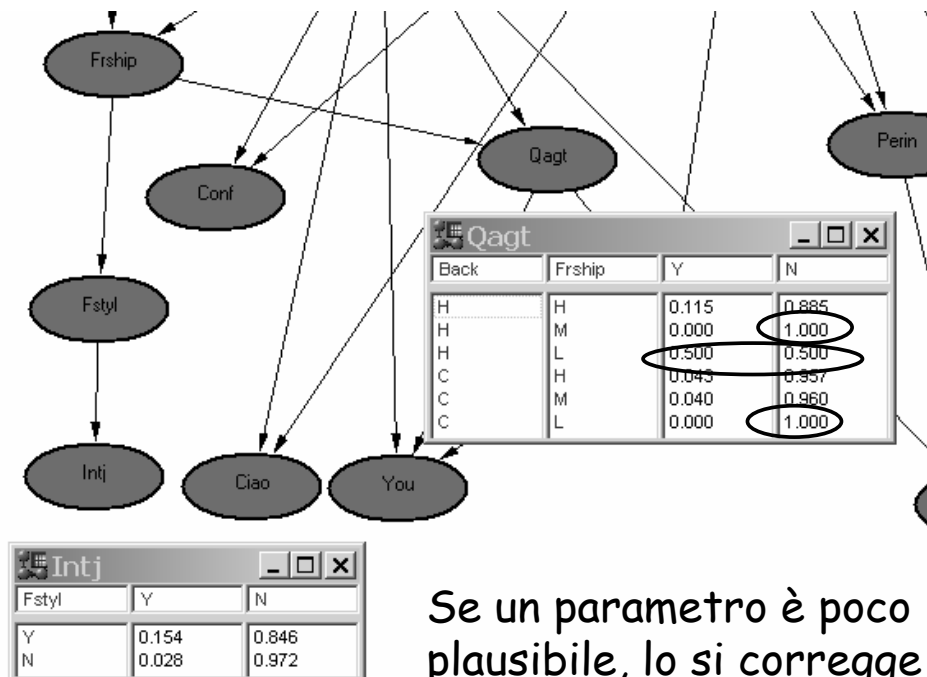


Quesito

Quali delle variabili nella RCP precedente
sono 'hidden'
e quali sono 'observable'?

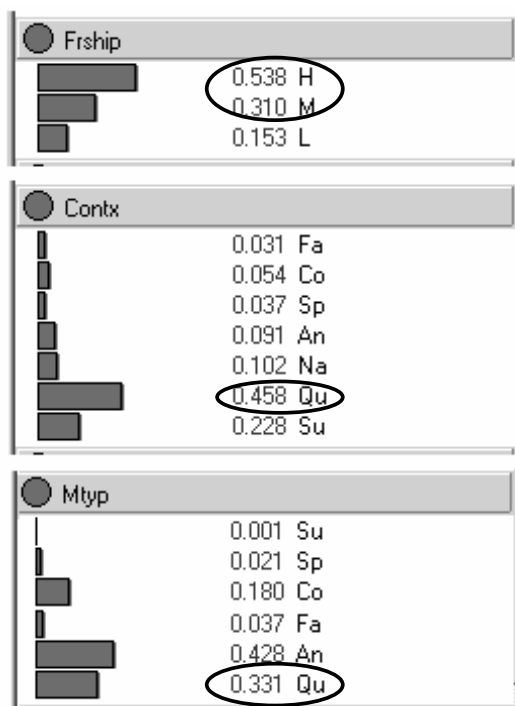
Apprendimento di RCP

1. Verificare la plausibilità dei parametri calcolati dall'algoritmo



Apprendimento di RCP

2. Verificare la plausibilità delle probabilità a priori dei nodi calcolate dall'algoritmo



Questo è un metodo efficace per verificare se i parametri introdotti sono corretti ed eventualmente rivederli.

È plausibile che il sistema faccia domande Nel 50% dei casi (Contx) e l'utente nel 33 % dei casi (MTyp)?

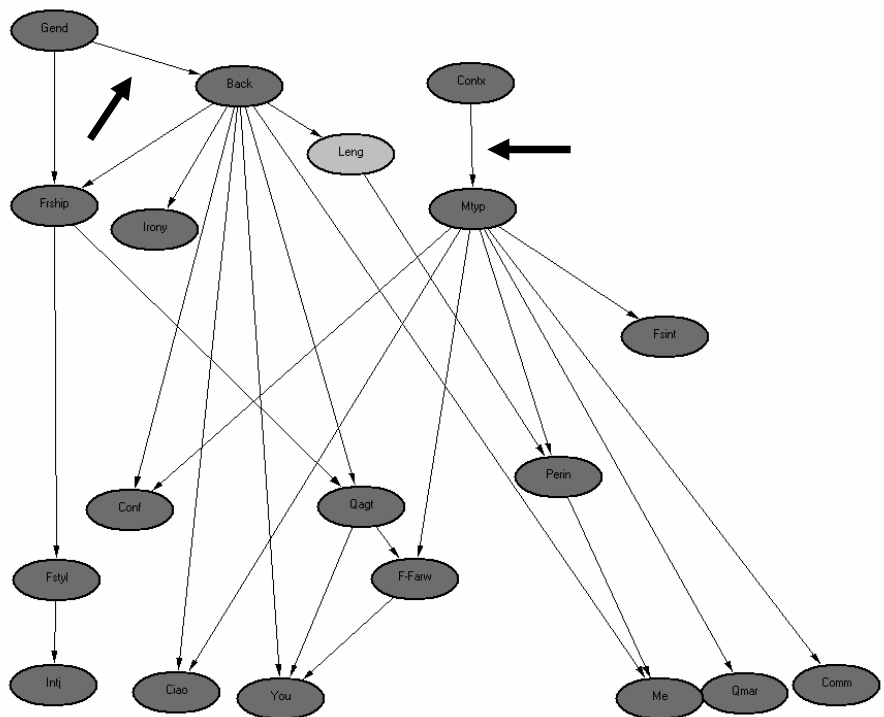
E' plausibile la distribuzione dell'atteggiamento dell'utente (Frship), nell'85% dei casi alta o media?

Apprendimento di RCP

3. Verificare la plausibilità di link 'appresi'

E' plausibile che il sesso (Gend) influenzi il background culturale (Back)?

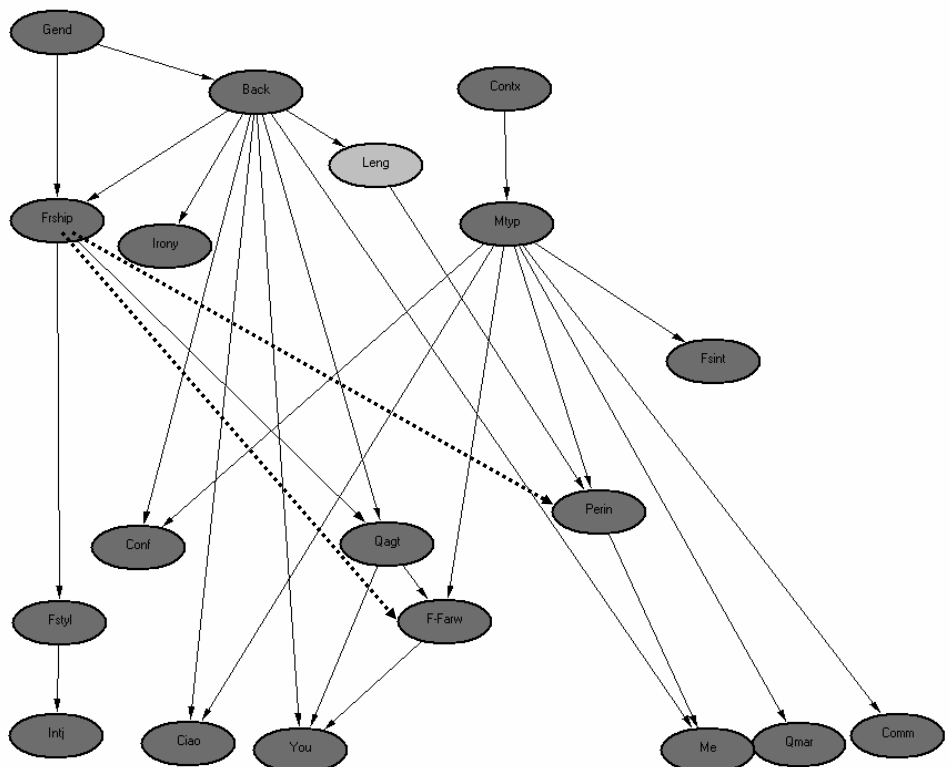
O che il tipo di 'move' dell'utente (Mtyp) dipenda dalla move effettuata dal sistema (Contx)?



Apprendimento di RCP

4. Verificare la mancanza di link 'attesi'

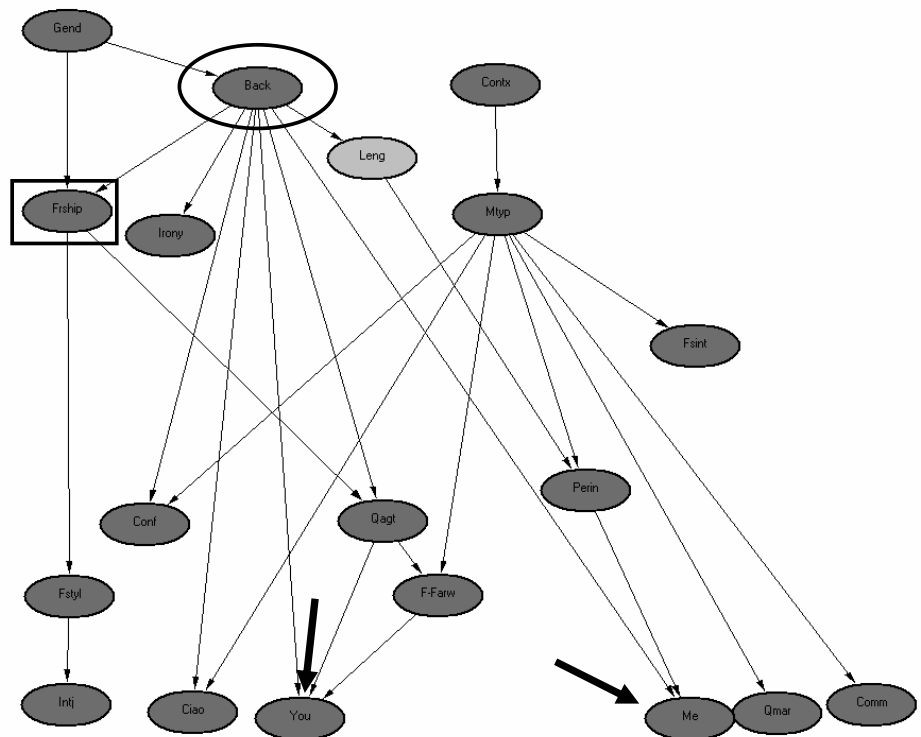
E' accettabile che manchi un link che lega l'atteggiamento 'amichevole' (Frship) al parlare di sé (Perin) o a salutare amichevolmente?



Apprendimento di RCP

5. Verificare l'effetto della 'lontananza' fra nodi

In questa rete, parlare di sé (Me) e fare domande sull'agente (You) influenzano la probabilità che l'utente abbia un atteggiamento 'amichevole' (Frship), ma debolmente, vista la lunghezza dei cammini che legano queste coppie di nodi.



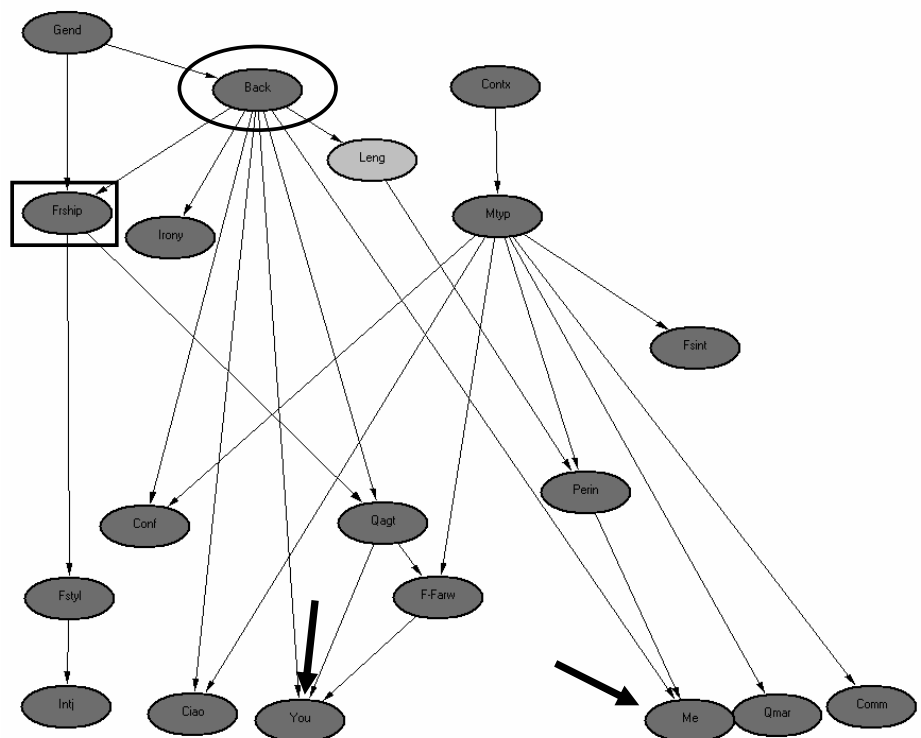
Apprendimento di RCP

5. Verificare l'effetto della 'lontananza' fra nodi

Pongo:
Gender = M
Back = H
Ottengo
 $P(\text{Frship}) = .792$

Inserisco:
You = Yes
ottengo
 $P(\text{Frship}) = .878$
(cammino lungo 2)

Se, invece, inserisco:
Me = Yes
ottengo
 $P(\text{Frship}) = .793$
(cammino lungo 5)



Apprendimento di RCP

6. Verificare l'effetto della proprietà di d-separazione

Consideriamo una RCP che contenga 5 nodi: A, B, C, D, E.

La sua probabilità congiunta (o *marginale*) sarà in genere uguale (per il Teorema di Bayes) a:

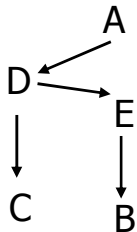
$$P(A, B, C, D, E) = P(A | B, C, D, E) * P(B | C, D, E) * P(C | D, E) * P(D | E) * P(E)$$

... e idem per i diversi valori delle 5 variabili...

In alcuni casi, si può assumere che siano verificate delle ipotesi di *indipendenza condizionale* (o d-separazione) fra alcune delle variabili, che semplificano il calcolo delle prob marginali.

Ad esempio, se B è indipendente da C e D dato E, avremo che:

$$P(B | C, D, E) = P(B | E)$$



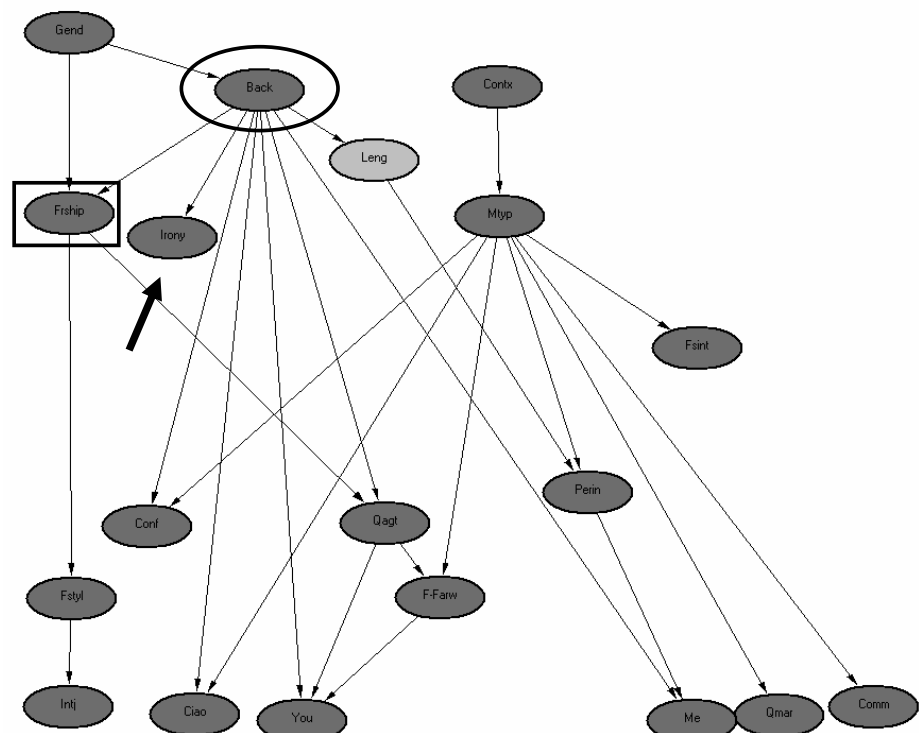
Questo significa che, se introduco una 'evidenza' su E, (ad es, E è 'vero'), la probabilità di B varierà. Ma a questo punto, introdurre evidenza su A, D o C non avrà più nessun effetto su B!!

Apprendimento di RCP

6. Verificare l'effetto della proprietà di d-separazione

Back d-separa Irony dal resto della rete.

Cioè:
una volta fissato il valore di Back, l'evidenza su Irony non si propaga su nessun altro nodo (ad esempio, Frship)



Apprendimento di RCP

6. Verificare l'effetto della proprietà di d-separazione

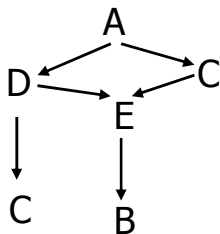
Nota che, nell'esempio, anche:

A è d-separato da E dato D

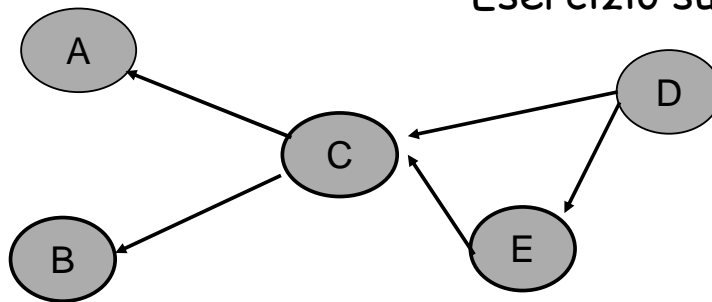
C è d-separato da B dato D

C è d-separato da B dato E

...



Esercizio sulla d-separazione



a) Which of the following statements are implied by the indicated network structure; answer yes and no; and give a brief reason for your answer!

i) $P(A,B|C) = P(A|C) \cdot P(B|C)$

yes, because...

ii) $P(C,E|D) = P(C|D) \cdot P(E|D)$

no, because...

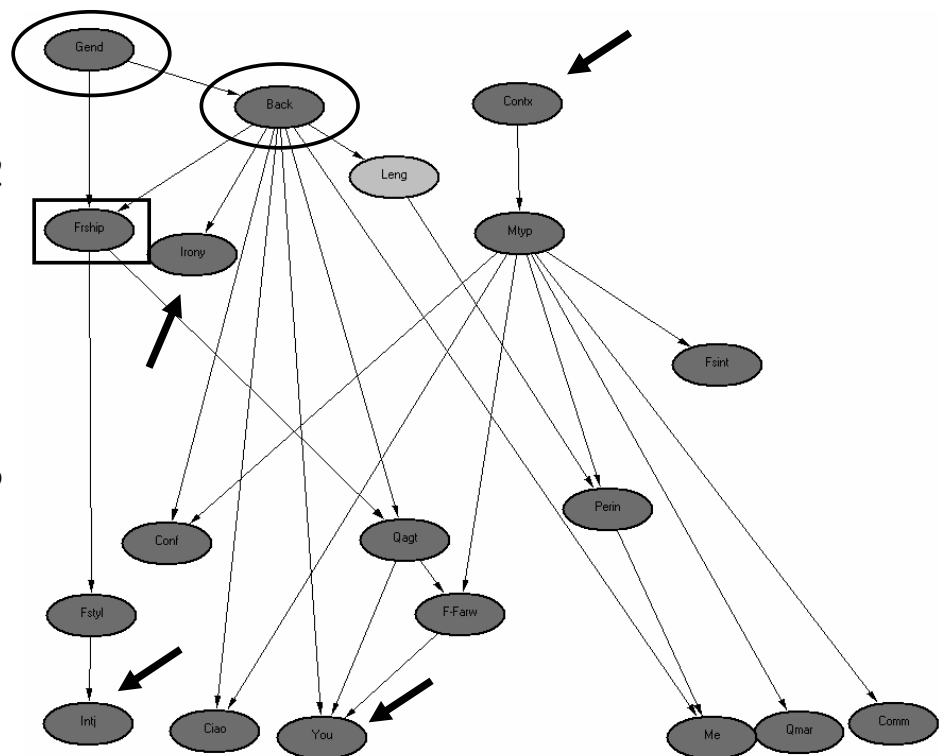
iii) $P(C|A) = P(C)$

no, because...

Apprendimento di RCP

6. Verificare l'effetto della proprietà di d-separazione

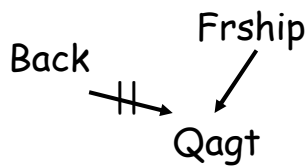
Back = H
Gender = M
→
 $P(\text{Frship} = H) = .792$
Contx = Qu
Intj = Y
→
 $P(\text{Frship} = H) = .821$
You = Y
 $P(\text{Frship} = H) = .906$
Irony = Y
 $P(\text{Frship} = H)$ non
cambia!



Apprendimento di RCP

Come correggere eventuali errori rilevati nel modello?

1. Errori sui parametri:
facili da correggere, in modo iterativo
2. Errori sui link:
si possono aggiungere o togliere dei link, ma *bisogna rivedere i parametri*. Processo iterativo.
3. Errori dovuti alla d-separazione:
sono i più complessi da gestire.



Come correggere i parametri quando si cancella un link


Back	Frship	Y	N
H	H	0.115	0.885
H	M	0.000	1.000
H	L	0.500	0.500
C	H	0.043	0.957
C	M	0.040	0.960
C	L	0.000	1.000

Obiettivo:

conservare la relazione fra il nodo considerato e i genitori non cancellati

Metodo:

media dei valori corrispondenti



Frship	Y	N
H	.079	.921
M	.020	.980
L	.250	.750

Come correggere problemi di d-separazione: Apprendimento Semi-Automatico

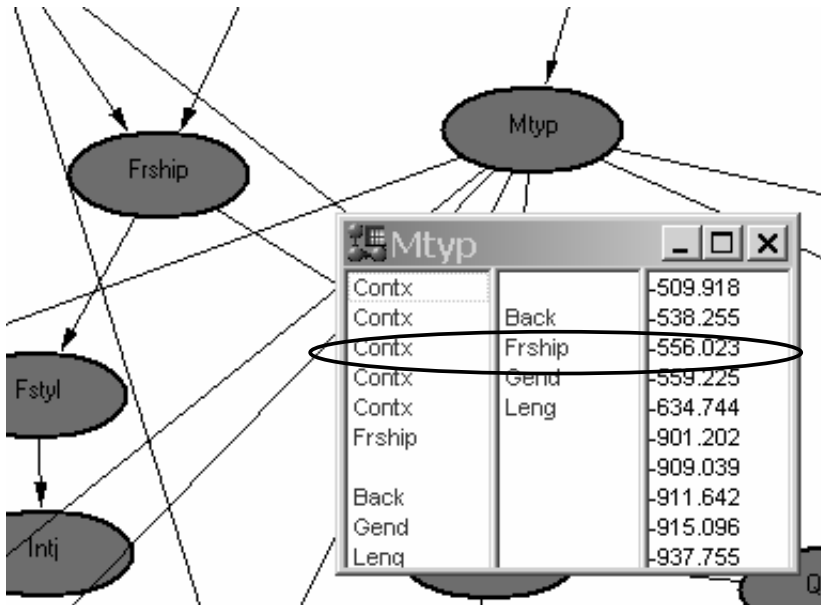
E' possibile imporre vincoli sulla struttura della rete,
e lasciare all'algoritmo il compito di completarla.

L'algoritmo aggiungerà dei link a quelli imposti,
calcolando i valori di tutti i parametri.

Il rischio è di ottenere una struttura *molto complessa*.

Apprendimento Semi-Automatico di RCP

Esempio: per correggere il problema di d-separazione visto prima, Potremmo inserire un link tra Mtyp e Frship.

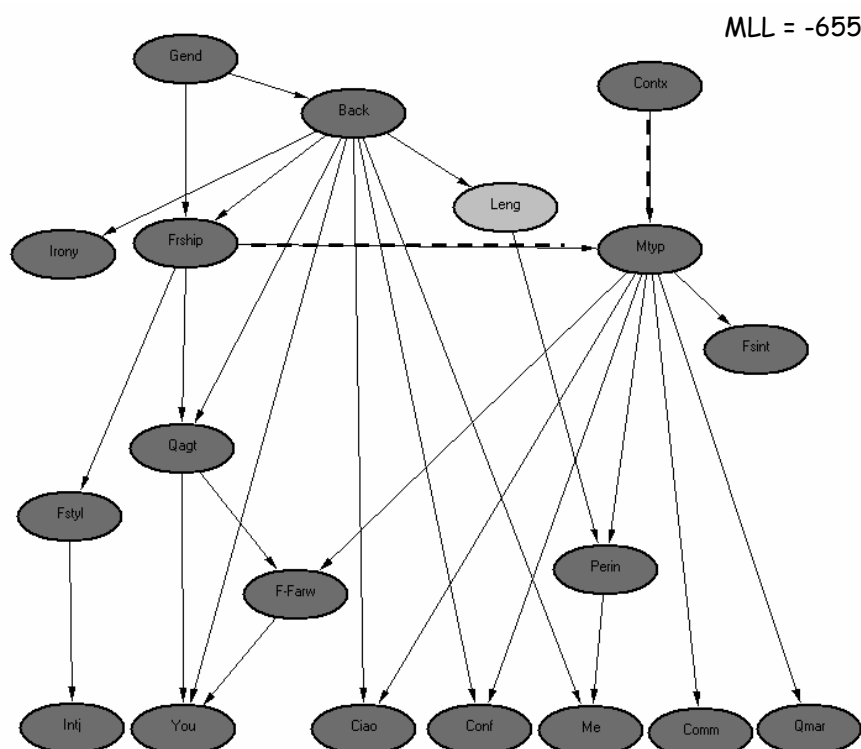


Mi chiedo allora:
l'algoritmo aveva
considerato questa ipotesi,
nella costruzione della
rete?

Che livello di MLL aveva
ottenuto?

Provo ad imporre questo
vincolo...

Apprendimento Semi-Automatico di RCP



La MLL è leggermente
diminuita,

il resto della
struttura è rimasto
invariato,

così come le
prior dei nodi.

Il problema
della d-separazione è
attenuato.

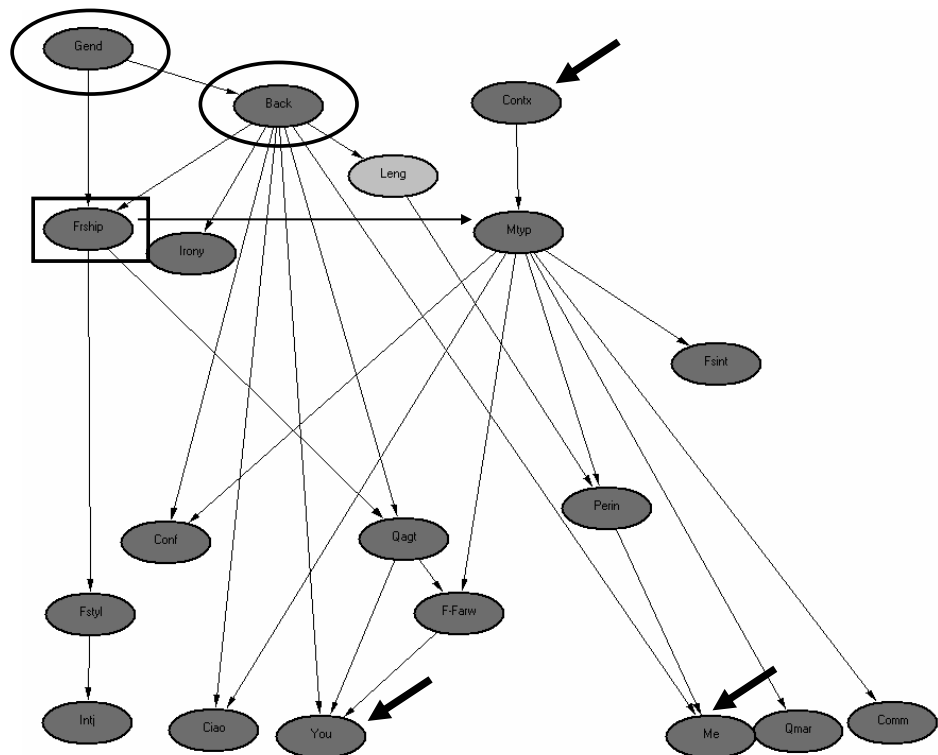
Apprendimento di RCP

Back = H
Gender = M
→
 $P(\text{Frship} = H) = .792$
Contx = Qu

You = Y
 $P(\text{Frship} = H) = .885$
Me = Y
 $P(\text{Frship} = H) = .890$

**Ma non ho risolto il
Problema di Irony!!**

Perché?????



Test del Modello

Ci sono due modi per testare la bontà di un modello:

- Il modo più classico (*test interno*) consiste nel dividere l'insieme D dei dati, in modo casuale, in un *training set*, (che comprenda ad es, il 75% dei casi) e in un *test set* (che comprenda il restante 25 %), di 'costruire' il modello sul training set e di verificare poi la accuratezza delle previsioni sul test set.
- Un modo più pragmatico (*test esterno*) addestra il modello su tutti i dati di D e lo sottopone poi a valutazione verificandone il comportamento su un *nuovo gruppo di casi reali* (ad es, di utenti).

Quali Tool sono disponibili per l'apprendimento automatico di RCP?

- Hugin
- Bayesware
- BNJ
- Netica
-

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