Intelligenza Artificiale II

Ragionamento probabilistico

Inferenza

Marco Piastra

Parte 2

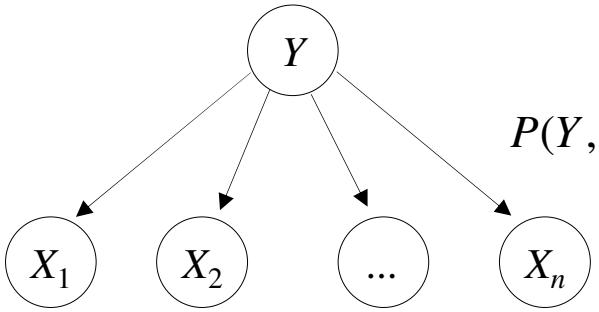
Problemi di inferenza probabilistica
Inferenza nei modelli grafici
Eliminazione delle variabili
Factor graphs e *belief propagation*

Modelli grafici ed inferenza

- Cosa si intende
  - A partire da un modello grafico completamente specificato (struttura + numeri)
    - Quindi da una distribuzione congiunta di probabilità
  - Il calcolo del valore di probabilità condizionale di alcuni nodi
    - Intesi come 'risposte', dato il valore di altri nodi, intesi come 'fatti osservati'
  - Con eventuale marginalizzazione
    - Si ignora il valore di nodi non osservati e non rilevanti come risposte

Esempio: filtro *anti-spam*

Tipicamente si usa un '*Naive* (*Discrete*) *Bayesian Classifier*'



$$P(Y, X_1, X_2, \ldots, X_n) = P(Y) \prod_{i=1}^{n} P(X_i \mid Y)$$

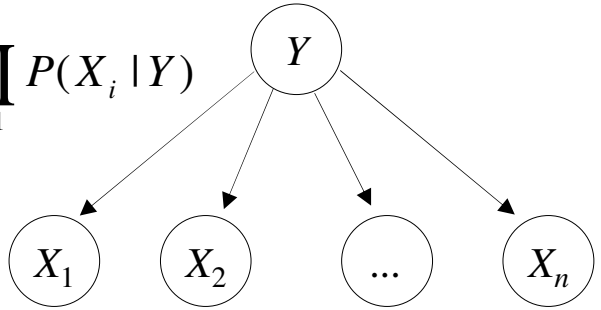
Nel caso del filtro *anti-spam*:

- Le variabili aleatorie sono tutte *binomiali* (valore 0 o 1)
- Y rappresenta la classe di appartenenza di un'email: 1 *spam*, 0 non-*spam*
- Ciascuna X_i indica la presenza nel messaggio di una determinata parola

Le probabilità associate al modello grafico si assumono come date

La determinazione delle probabilità condizionali, tipicamente partendo da dati campione, è un problema di *apprendimento* (vedi oltre)

$$P(Y, X_1, X_2, \ldots, X_n) = P(Y) \prod_{i=1}^{n} P(X_i \mid Y)$$



Inferenza nel filtro *anti-spam*

Data una combinazione di parole $\{X_k\}$, si cerca la classificazione a maggiore probabilità

L'email è *spam* se $\dfrac{P(Y=1 \mid \{X_k\})}{P(Y=0 \mid \{X_k\})} > \lambda$

Notare che:

$$P(Y=1 \mid \{X_k\}) = \frac{P(\{X_k\} \mid Y=1)P(Y=1)}{\sum_Y P(\{X_k\} \mid Y)P(Y)} = \frac{P(Y=1)\prod_k P(X_k \mid Y=1)}{\sum_Y P(Y)\prod_k P(X_k \mid Y)}$$

Teorema di Bayes

Indipendenza condizionale, dal grafo

Quindi:

$$\frac{P(Y=1 \mid \{X_k\})}{P(Y=0 \mid \{X_k\})} = \frac{P(Y=1)}{P(Y=0)} \prod_k \frac{P(X_k \mid Y=1)}{P(X_k \mid Y=0)}$$

Per semplificare i calcoli si usa il logaritmo:

$$\log \frac{P(Y=1 \mid \{X_k\})}{P(Y=0 \mid \{X_k\})} = \log \frac{P(Y=1)}{P(Y=0)} + \sum_k \log \frac{P(X_k \mid Y=1)}{P(X_k \mid Y=0)}$$

Inferenza probabilistica

- In generale

  Si assume come punto di partenza una distribuzione congiunta completamente specificata

  $P(X_1, X_2, .., X_n)$

  In un problema di inferenza probabilistica, le variabili aleatorie $\{X_1, X_2, ..., X_n\}$ si dividono in tre categorie:

  1) Variabili $\{X_e\}$ *osservate* e che quindi hanno un valore definito
  2) Variabili $\{X_r\}$ non rilevanti, per il problema in esame
  3) Variabili $\{X_f\}$ che rappresentano la risposta al problema in esame

  Si tratta di trovare: $P(\{X_f\}|\{X_e\}) = \sum_{\{X_r\}} P(\{X_f\},\{X_r\}|\{X_e\})$

  - Non esiste un problema di decidibilità: il calcolo diretto è sempre fattibile (*nel discreto)

    La distribuzione di probabilità è completamente specificata
  - Esiste semmai un problema di efficienza

    La quantità di numeri da trattare cresce esponenzialmente con il numero delle variabili e l'estensione del dominio di queste

Costruzione di un modello grafico

- Passo 1

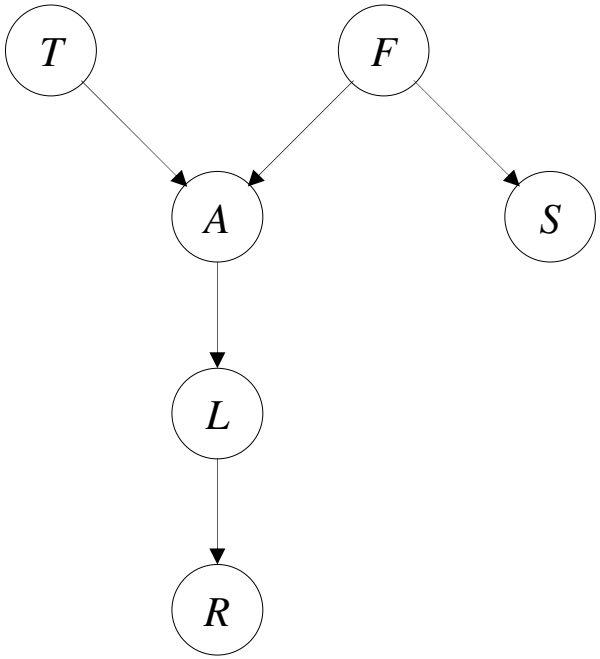
  Individuazione delle variabili

T : (*tampering*) – Manomissione del sensore
F : (*fire*) – Presenza di incendio
A : (*alarm*) – Allarme
S : (*smoke*) – Presenza di fumo
L : (*leaving*) – Abbandono della zona
R : (*report*) – Chiamata dei vigili del fuoco

Costruzione di un modello grafico

- Passo 2

  Definizione della struttura

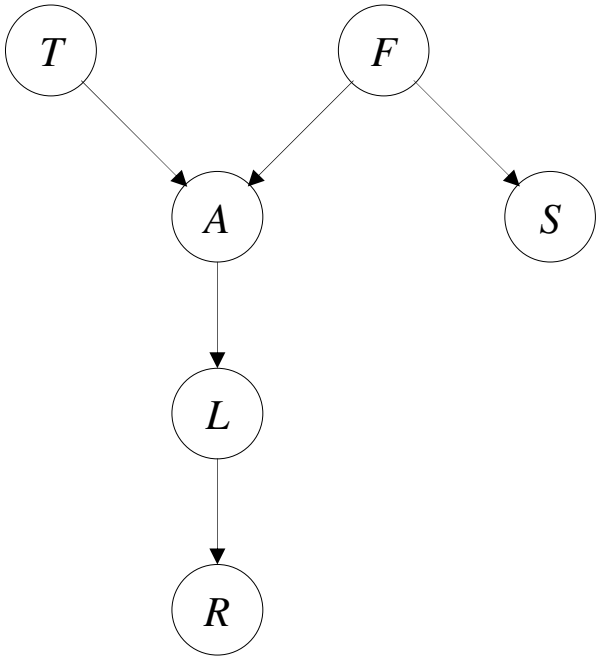


T : (*tampering*) – Manomissione del sensore
F : (*fire*) – Presenza di incendio
A : (*alarm*) – Allarme
S : (*smoke*) – Presenza di fumo
L : (*leaving*) – Abbandono della zona
R : (*report*) – Chiamata dei vigili del fuoco

Costruzione di un modello grafico

- Passo 2

  Definizione della struttura



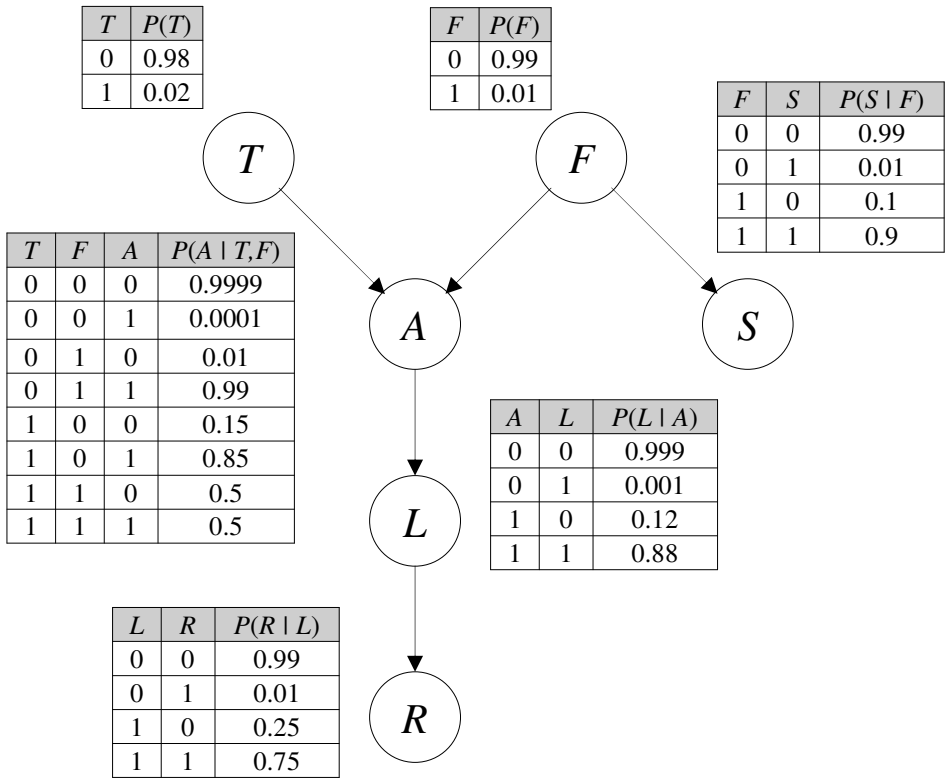
In base al modello grafico:
$<T \perp F>$ (ma diventano dipendenti se una delle di A, L, R, è noto)
$<A \perp S \mid F>$
$<L \perp T \mid A>$
$<L \perp F \mid A>$
$<A \perp R \mid L>$

T : (*tampering*) – Manomissione del sensore
F : (*fire*) – Presenza di incendio
A : (*alarm*) – Allarme
S : (*smoke*) – Presenza di fumo
L : (*leaving*) – Abbandono della zona
R : (*report*) – Chiamata dei vigili del fuoco

Costruzione di un modello grafico

- Passo 3

  Definizione delle probabilità condizionali



T	$P(T)$
0	0.98
1	0.02

F	$P(F)$
0	0.99
1	0.01

F	S	$P(S \mid F)$
0	0	0.99
0	1	0.01
1	0	0.1
1	1	0.9

T	F	A	$P(A \mid T,F)$
0	0	0	0.9999
0	0	1	0.0001
0	1	0	0.01
0	1	1	0.99
1	0	0	0.15
1	0	1	0.85
1	1	0	0.5
1	1	1	0.5

A	L	$P(L \mid A)$
0	0	0.999
0	1	0.001
1	0	0.12
1	1	0.88

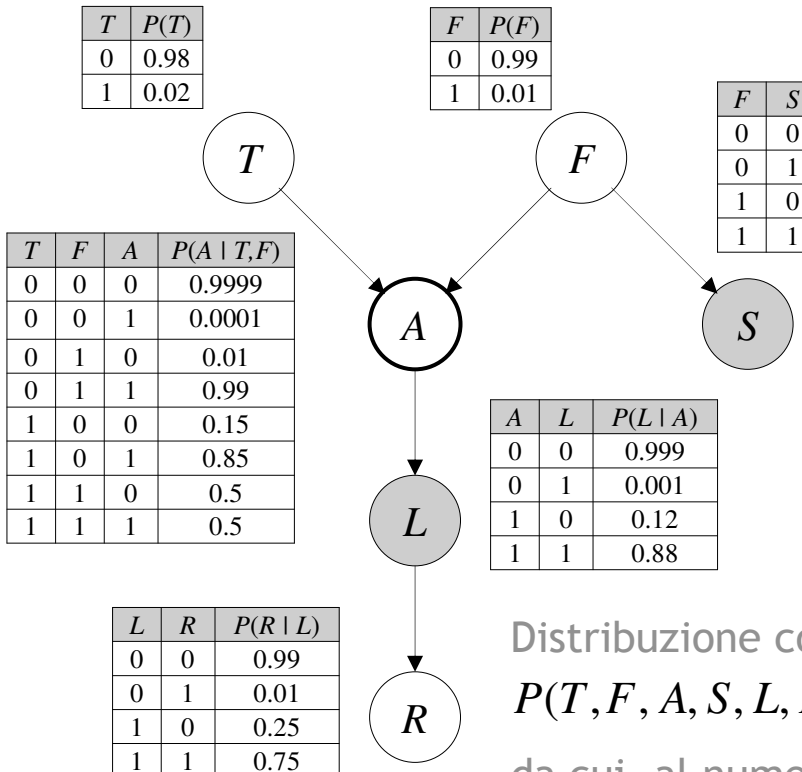
L	R	$P(R \mid L)$
0	0	0.99
0	1	0.01
1	0	0.25
1	1	0.75

T : (*tampering*) – Manomissione del sensore
F : (*fire*) – Presenza di incendio
A : (*alarm*) – Allarme
S : (*smoke*) – Presenza di fumo
L : (*leaving*) – Abbandono della zona
R : (*report*) – Chiamata dei vigili del fuoco

Inferenza probabilistica

- Passo 4

Impostazione di un problema



T	$P(T)$
0	0.98
1	0.02

F	$P(F)$
0	0.99
1	0.01

F	S	$P(S \mid F)$
0	0	0.99
0	1	0.01
1	0	0.1
1	1	0.9

T	F	A	$P(A \mid T,F)$
0	0	0	0.9999
0	0	1	0.0001
0	1	0	0.01
0	1	1	0.99
1	0	0	0.15
1	0	1	0.85
1	1	0	0.5
1	1	1	0.5

A	L	$P(L \mid A)$
0	0	0.999
0	1	0.001
1	0	0.12
1	1	0.88

L	R	$P(R \mid L)$
0	0	0.99
0	1	0.01
1	0	0.25
1	1	0.75

Ad esempio, calcolare la probabilità di A date le osservazioni $L=1$ e $S=0$

$$P(A \mid L=1, S=0) = \frac{P(A, L=1, S=0)}{P(L=1, S=0)}$$

Distribuzione congiunta, dal grafo:

$$P(T,F,A,S,L,R) = P(T)P(F)P(A \mid T,F)P(S \mid F)P(L \mid A)P(R \mid L)$$

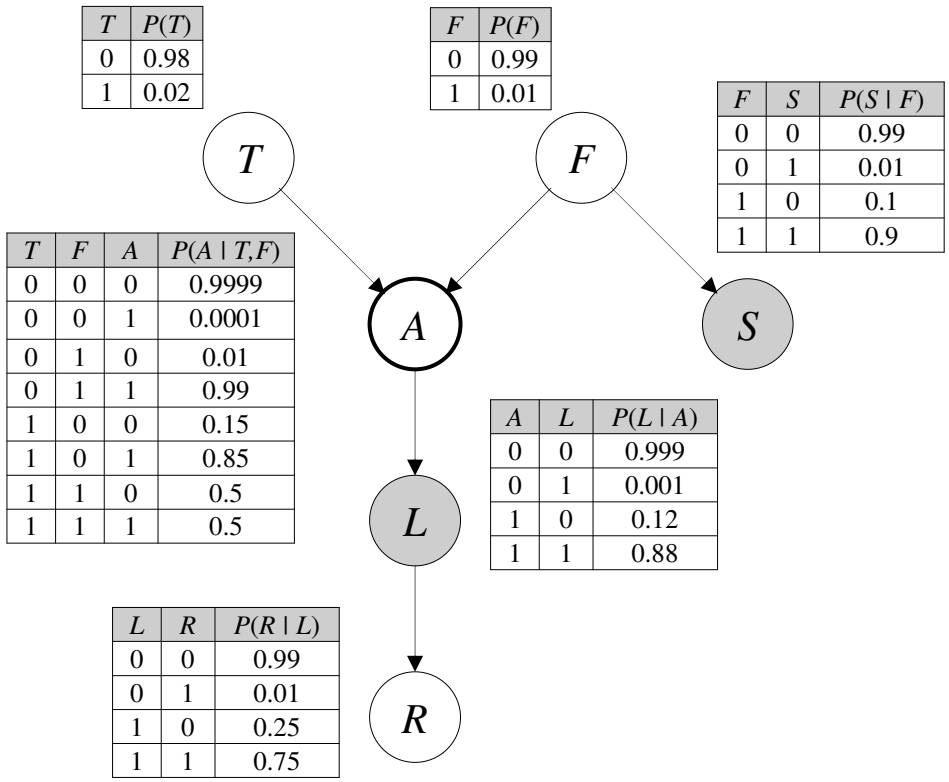
da cui, al numeratore:

$$P(A, L=1, S=0) = \sum_{T,F,R} P(T)P(F)P(A \mid T,F)P(S=0 \mid F)P(L=1 \mid A)P(R \mid L=1)$$

Inferenza probabilistica

- Passo 5

  Calcolo del problema



T	P(T)
0	0.98
1	0.02

F	P(F)
0	0.99
1	0.01

F	S	P(S | F)
0	0	0.99
0	1	0.01
1	0	0.1
1	1	0.9

T	F	A	P(A | T,F)
0	0	0	0.9999
0	0	1	0.0001
0	1	0	0.01
0	1	1	0.99
1	0	0	0.15
1	0	1	0.85
1	1	0	0.5
1	1	1	0.5

A	L	P(L | A)
0	0	0.999
0	1	0.001
1	0	0.12
1	1	0.88

L	R	P(R | L)
0	0	0.99
0	1	0.01
1	0	0.25
1	1	0.75

Notare che in:

$$P(A \mid L=1, S=0) = \frac{P(A, L=1, S=0)}{P(L=1, S=0)}$$

Questo termine serve a normalizzare: può essere calcolato a partire da

$$P(A, L=1, S=0)$$

Infatti:

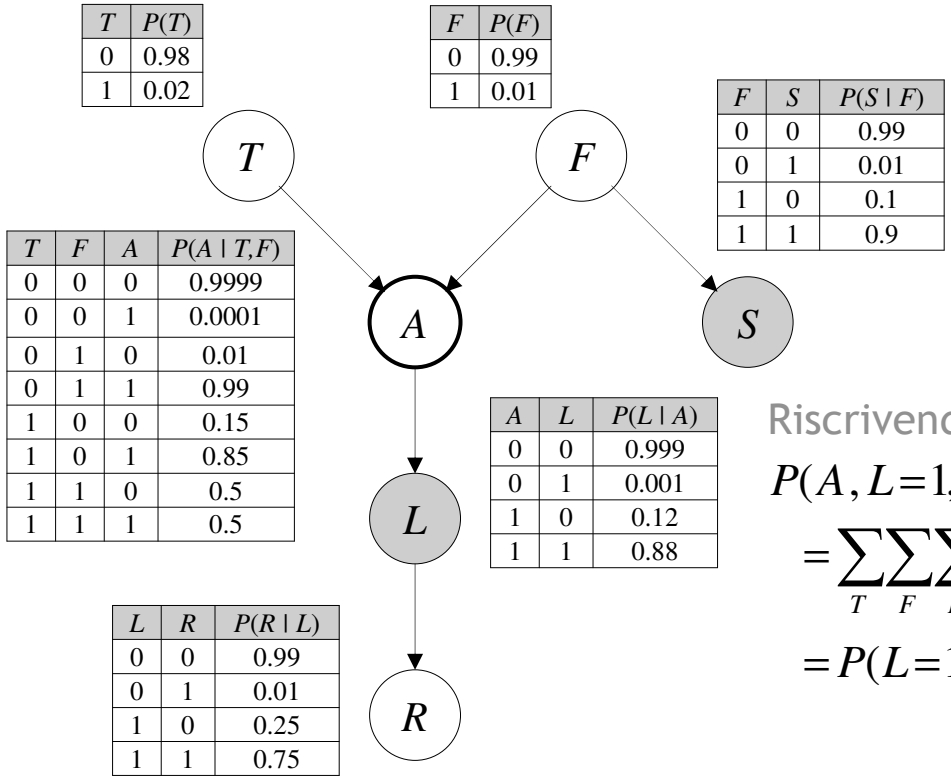
$$P(L=1, S=0) = \sum_{A} P(A, L=1, S=0)$$

Tipicamente, la parte dominante del calcolo di un problema di inferenza probabilistica è la marginalizzazione

Inferenza probabilistica

- Passo 5

  Calcolo del problema



T	$P(T)$
0	0.98
1	0.02

F	$P(F)$
0	0.99
1	0.01

F	S	$P(S \mid F)$
0	0	0.99
0	1	0.01
1	0	0.1
1	1	0.9

T	F	A	$P(A \mid T,F)$
0	0	0	0.9999
0	0	1	0.0001
0	1	0	0.01
0	1	1	0.99
1	0	0	0.15
1	0	1	0.85
1	1	0	0.5
1	1	1	0.5

A	L	$P(L \mid A)$
0	0	0.999
0	1	0.001
1	0	0.12
1	1	0.88

L	R	$P(R \mid L)$
0	0	0.99
0	1	0.01
1	0	0.25
1	1	0.75

Riscrivendo e utilizzando la proprietà distributiva:

$$P(A, L=1, S=0)$$

$$= \sum_T \sum_F \sum_R P(L=1 \mid A) P(A \mid T, F) P(T) P(F) P(S=0 \mid F) P(R \mid L=1)$$

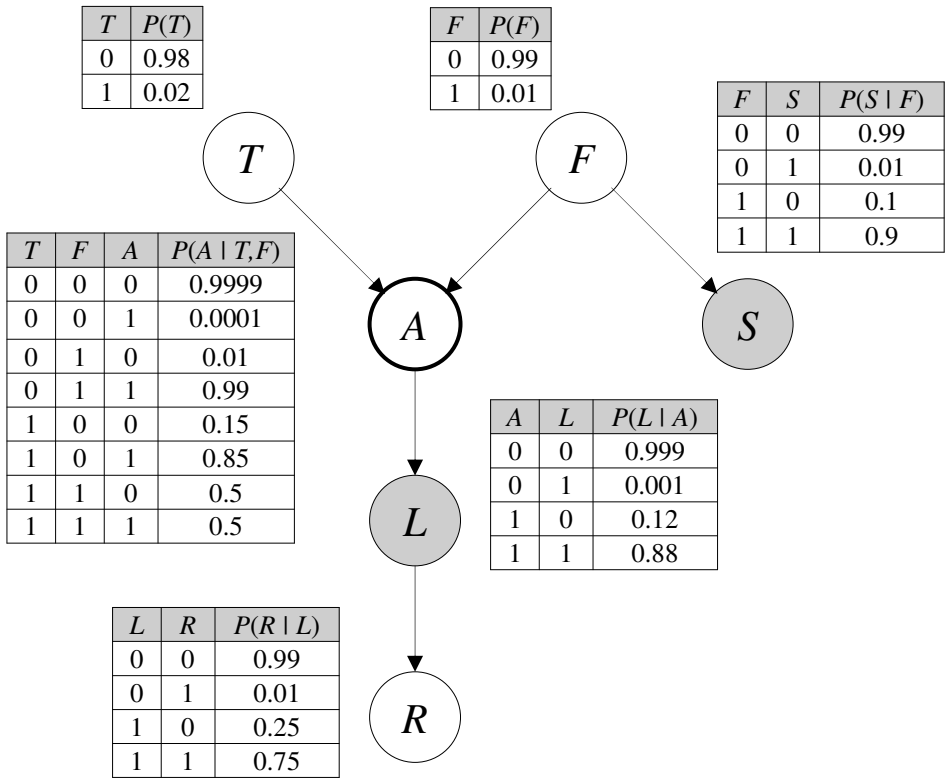
$$= P(L=1 \mid A) \sum_T \sum_F P(A \mid T, F) P(T) P(F) P(S=0 \mid F) \sum_R P(R \mid L=1)$$

Questa sommatoria vale 1
C'era da aspettarselo
in quanto $<A \perp R \mid L>$

Inferenza probabilistica

- Passo 5

  Calcolo del problema



T	$P(T)$
0	0.98
1	0.02

F	$P(F)$
0	0.99
1	0.01

F	S	$P(S \mid F)$
0	0	0.99
0	1	0.01
1	0	0.1
1	1	0.9

T	F	A	$P(A \mid T,F)$
0	0	0	0.9999
0	0	1	0.0001
0	1	0	0.01
0	1	1	0.99
1	0	0	0.15
1	0	1	0.85
1	1	0	0.5
1	1	1	0.5

A	L	$P(L \mid A)$
0	0	0.999
0	1	0.001
1	0	0.12
1	1	0.88

L	R	$P(R \mid L)$
0	0	0.99
0	1	0.01
1	0	0.25
1	1	0.75

Per convenzione, si scrive:

$$P(A, L=1, S=0) = f_{T,F,S=0}(A) f_{L=1}(A)$$

dove gli f sono i *fattori* del metodo di calcolo detto anche *eliminazione delle variabili*.

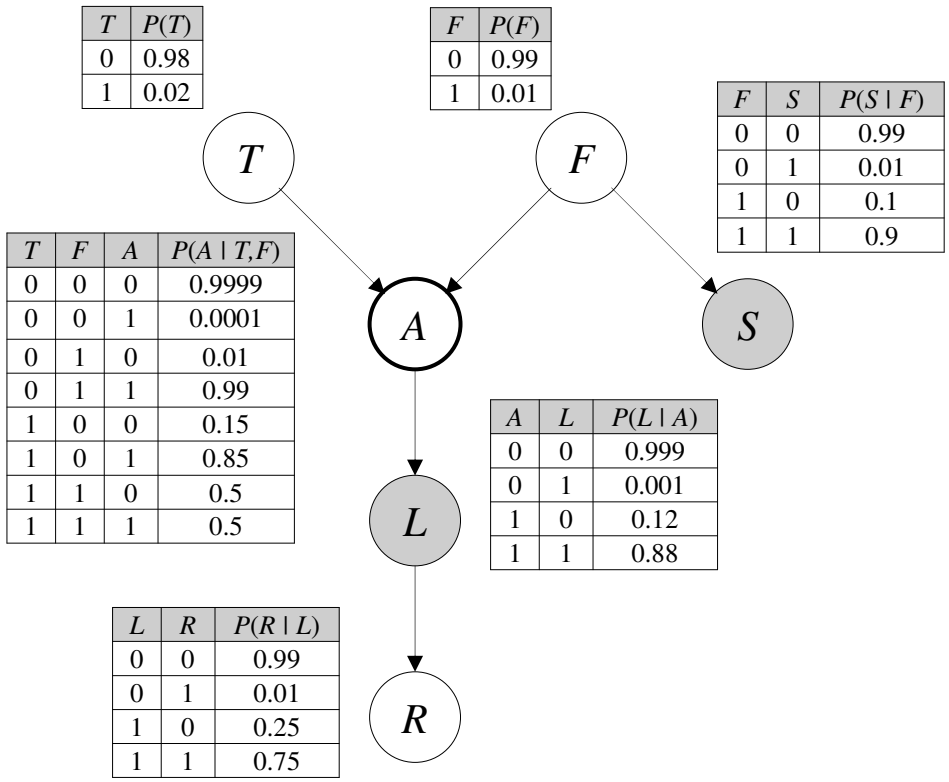
In generale i *fattori* f NON sono probabilità (non hanno somma 1). Ad esempio, non lo è:

$$f_{T,F,S=0}(A) = \sum_T \sum_F P(A \mid T, F) P(T) P(F) P(S=0 \mid F)$$

Inferenza probabilistica

- Passo 5

  Calcolo del problema



T	P(T)
0	0.98
1	0.02

F	P(F)
0	0.99
1	0.01

F	S	P(S | F)
0	0	0.99
0	1	0.01
1	0	0.1
1	1	0.9

T	F	A	P(A | T,F)
0	0	0	0.9999
0	0	1	0.0001
0	1	0	0.01
0	1	1	0.99
1	0	0	0.15
1	0	1	0.85
1	1	0	0.5
1	1	1	0.5

A	L	P(L | A)
0	0	0.999
0	1	0.001
1	0	0.12
1	1	0.88

L	R	P(R | L)
0	0	0.99
0	1	0.01
1	0	0.25
1	1	0.75

Per convenzione, si scrive:

$$P(A, L=1, S=0) = f_{T,F,S=0}(A) f_{L=1}(A)$$

dove gli f sono i *fattori* del metodo di calcolo detto anche *eliminazione delle variabili*.

In generale i *fattori* f NON sono probabilità (non hanno somma 1). Ad esempio, non lo è:

$$f_{T,F,S=0}(A) = \sum_T \sum_F P(A|T,F)P(T)P(F)P(S=0|F)$$

Motivo:
Sommando rispetto ad una variabile condizionata, si ha una probabilità marginale

$$P(A|S) = \sum_L P(A, L|S)$$

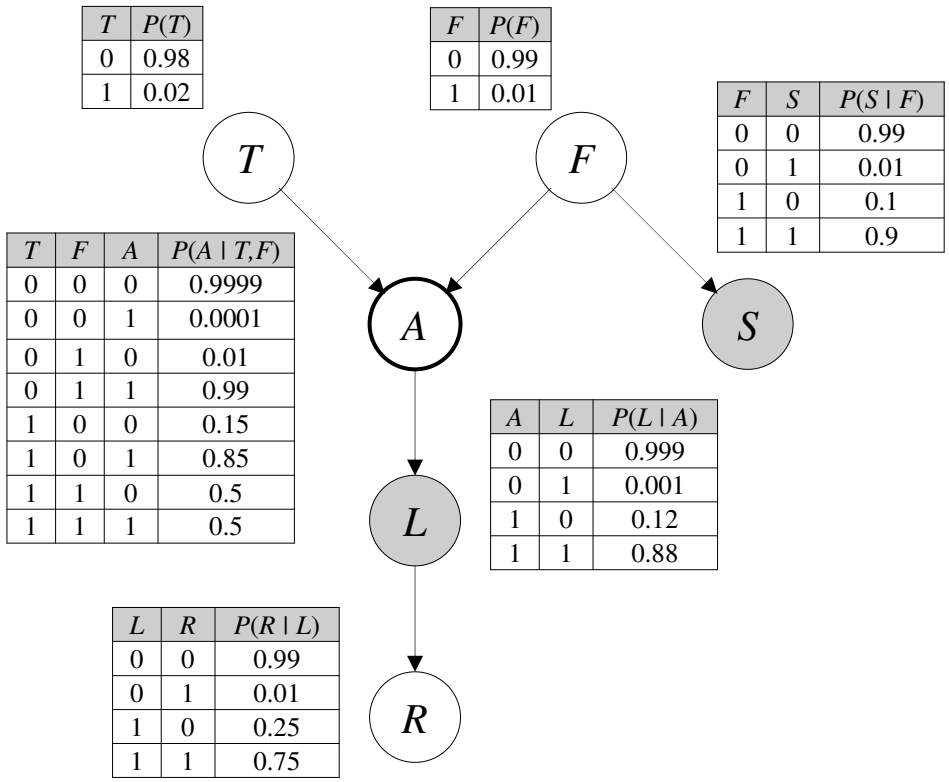
Sommando rispetto ad una variabile condizionante, si ha solo una funzione

$$f_S(A,L) = \sum_S P(A, L|S)$$

Inferenza probabilistica

- Passo 5

  Calcolo del problema



T	P(T)
0	0.98
1	0.02

F	P(F)
0	0.99
1	0.01

F	S	P(S | F)
0	0	0.99
0	1	0.01
1	0	0.1
1	1	0.9

T	F	A	P(A | T,F)
0	0	0	0.9999
0	0	1	0.0001
0	1	0	0.01
0	1	1	0.99
1	0	0	0.15
1	0	1	0.85
1	1	0	0.5
1	1	1	0.5

A	L	P(L | A)
0	0	0.999
0	1	0.001
1	0	0.12
1	1	0.88

L	R	P(R | L)
0	0	0.99
0	1	0.01
1	0	0.25
1	1	0.75

Notare:

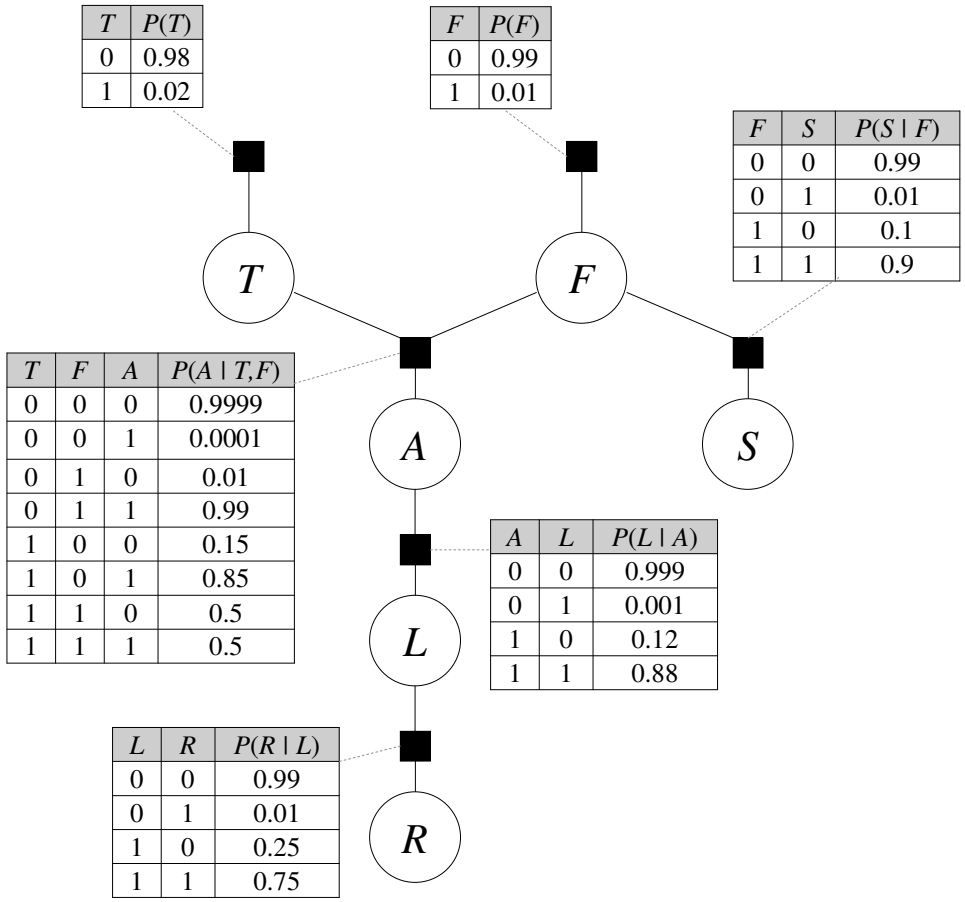
$$P(A, L=1, S=0) = f_{T,F,S=0}(A)\, f_{L=1}(A)$$

Un fattore proviene dai *parents* di A

Un fattore proviene dai *descendants* di A

Ciò è vero in generale, se A *d-separa* il grafo

Factor graphs



T	$P(T)$
0	0.98
1	0.02

F	$P(F)$
0	0.99
1	0.01

F	S	$P(S \mid F)$
0	0	0.99
0	1	0.01
1	0	0.1
1	1	0.9

T	F	A	$P(A \mid T,F)$
0	0	0	0.9999
0	0	1	0.0001
0	1	0	0.01
0	1	1	0.99
1	0	0	0.15
1	0	1	0.85
1	1	0	0.5
1	1	1	0.5

A	L	$P(L \mid A)$
0	0	0.999
0	1	0.001
1	0	0.12
1	1	0.88

L	R	$P(R \mid L)$
0	0	0.99
0	1	0.01
1	0	0.25
1	1	0.75

Un metodo di calcolo distribuito per l'inferenza probabilistica

Il modello grafico viene tradotto in un *grafo bipartito*:

a) Ai nodi che rappresentano le *variabili*

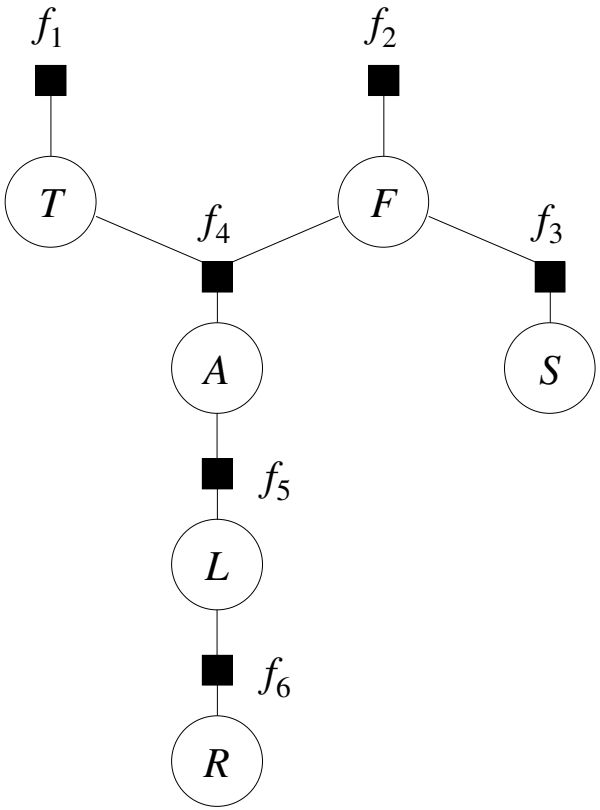
b) Si aggiungono i nodi che rappresentano le *funzioni*

Il grafo risultante non è più orientato

Un *factor graph* è un altro modo per rappresentare la fattorizzazione di una funzione complessiva delle variabili

In questo caso,
una distribuzione congiunta di probabilità

Factor graphs



Message passing (o *belief propagation*)

Si può immaginare che ciascun nodo di un *factor graph* sia un processore

Il calcolo delle probabilità marginali avviene tramite lo scambio di *messaggi*

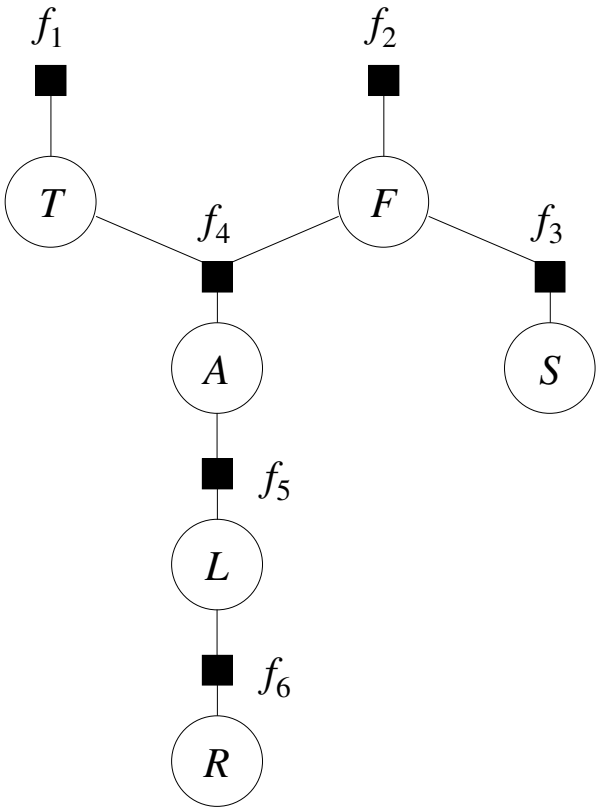
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)\setminus\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Message passing

Si può immaginare che ciascun nodo di un *factor graph* sia un processore

Il calcolo delle probabilità marginali avviene tramite lo scambio di *messaggi*

Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)\setminus\{f\}} \mu_{h\to x}(x)$$

Prodotto dei messaggi ricevuti dagli *altri* nodi funzione

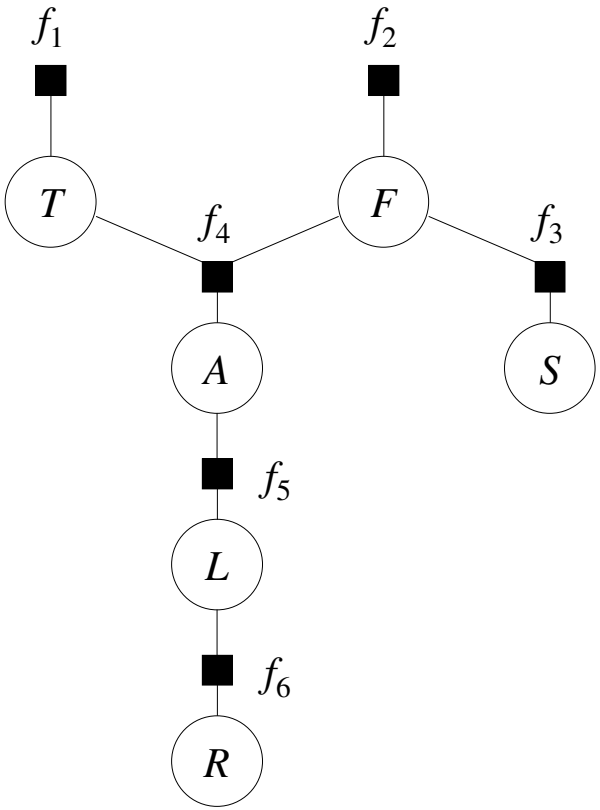
$\setminus\{f\}$ ad eccezione di f

$n(x)$ sono i vicini di x

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Message passing

Si può immaginare che ciascun nodo di un *factor graph* sia un processore

Il calcolo delle probabilità marginali avviene tramite lo scambio di *messaggi*

Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)\setminus\{f\}} \mu_{h\to x}(x)$$

Prodotto dei messaggi ricevuti dagli *altri* nodi funzione

$\setminus\{f\}$ ad eccezione di f

$n(x)$ sono i vicini di x

Da nodo funzione a nodo variabile:

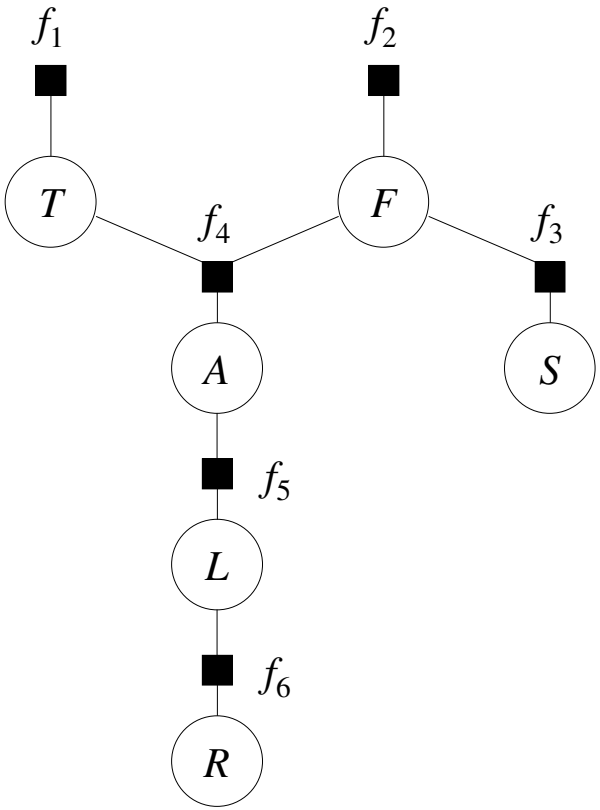
$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

X è il vettore delle variabili di f

Prodotto dei messaggi ricevuti dagli *altri* nodi variabile

Marginalizzazione rispetto a tutte le variabili connesse ad eccezione di x

Factor graphs



Message passing

Dalle formule si desume che:

- Ciascun nodo può inviare un messaggio ad un'altro nodo quando ha ricevuto i messaggi provenienti da tutti gli *altri* nodi
- I messaggi sono vettori (funzioni di una variabile), non scalari

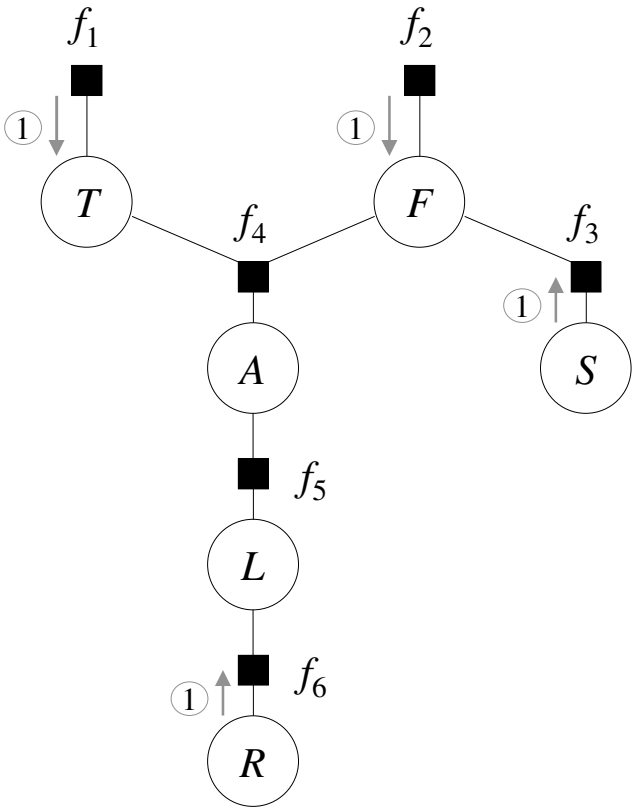
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)/\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Propagazione dei messaggi

Message passing

Dalle formule si desume che:

- Ciascun nodo può inviare un messaggio ad un'altro nodo quando ha ricevuto i messaggi provenienti da tutti gli *altri* nodi
- I messaggi sono vettori (funzioni di una variabile), non scalari

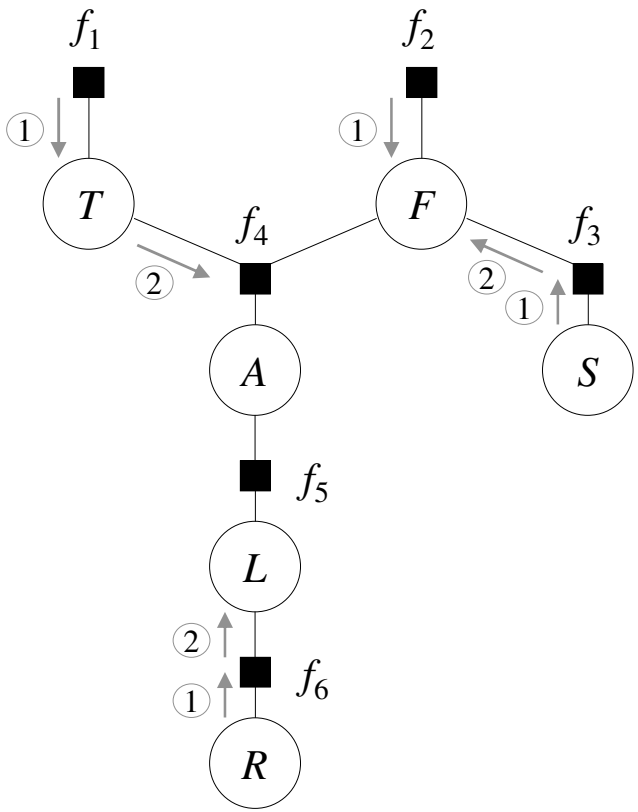
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)/\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Propagazione dei messaggi

Message passing

Dalle formule si desume che:

- Ciascun nodo può inviare un messaggio ad un'altro nodo quando ha ricevuto i messaggi provenienti da tutti gli *altri* nodi
- I messaggi sono vettori (funzioni di una variabile), non scalari

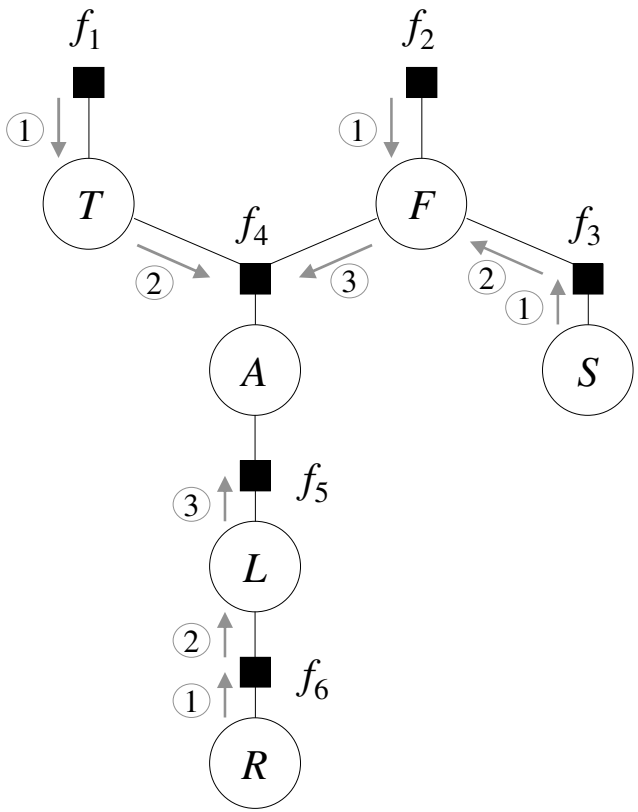
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)/\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Propagazione dei messaggi

Message passing

Dalle formule si desume che:

- Ciascun nodo può inviare un messaggio ad un'altro nodo quando ha ricevuto i messaggi provenienti da tutti gli *altri* nodi
- I messaggi sono vettori (funzioni di una variabile), non scalari

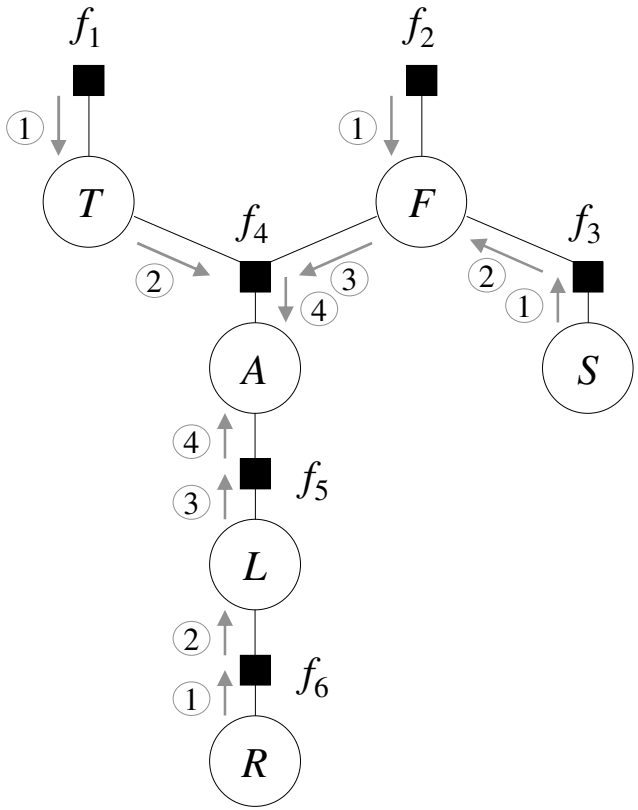
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)/\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Propagazione dei messaggi

Message passing

Dalle formule si desume che:

- Ciascun nodo può inviare un messaggio ad un'altro nodo quando ha ricevuto i messaggi provenienti da tutti gli *altri* nodi
- I messaggi sono vettori (funzioni di una variabile), non scalari

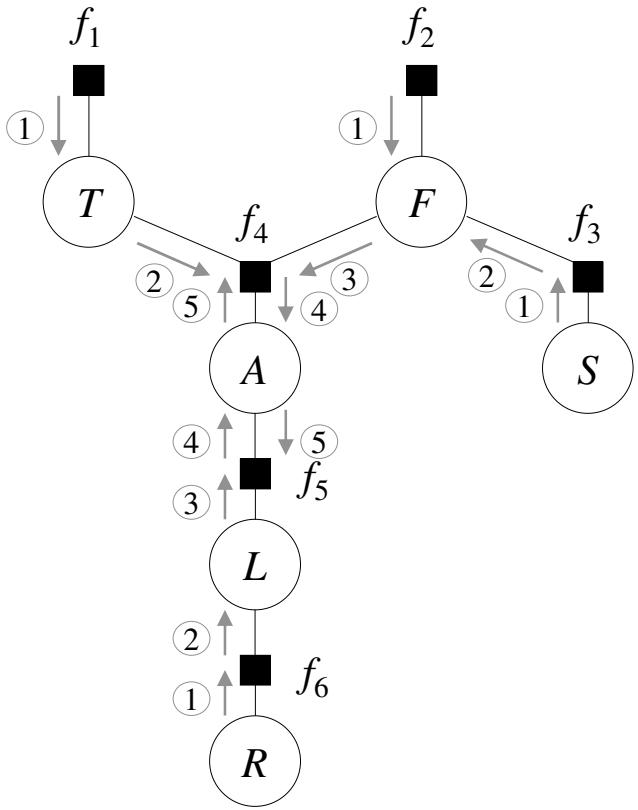
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)/\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Propagazione dei messaggi

Message passing

Dalle formule si desume che:

- Ciascun nodo può inviare un messaggio ad un'altro nodo quando ha ricevuto i messaggi provenienti da tutti gli *altri* nodi
- I messaggi sono vettori (funzioni di una variabile), non scalari

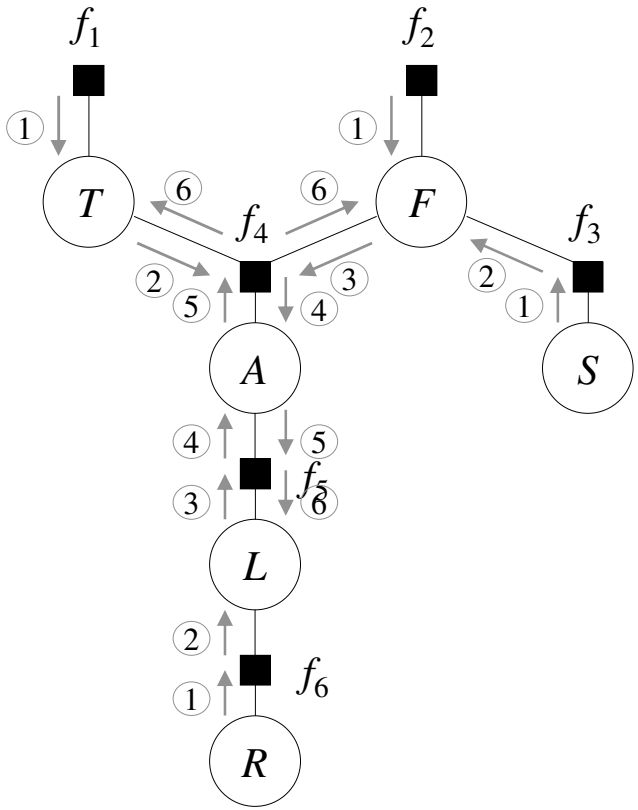
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)/\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Propagazione dei messaggi

Message passing

Dalle formule si desume che:

- Ciascun nodo può inviare un messaggio ad un'altro nodo quando ha ricevuto i messaggi provenienti da tutti gli *altri* nodi
- I messaggi sono vettori (funzioni di una variabile), non scalari

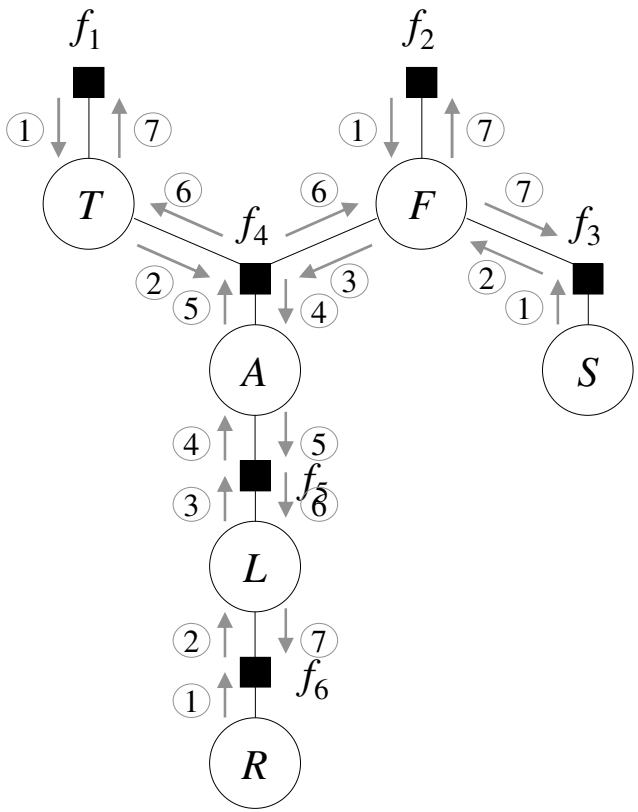
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)/\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Propagazione dei messaggi

Message passing

Dalle formule si desume che:

- Ciascun nodo può inviare un messaggio ad un'altro nodo quando ha ricevuto i messaggi provenienti da tutti gli *altri* nodi
- I messaggi sono vettori (funzioni di una variabile), non scalari

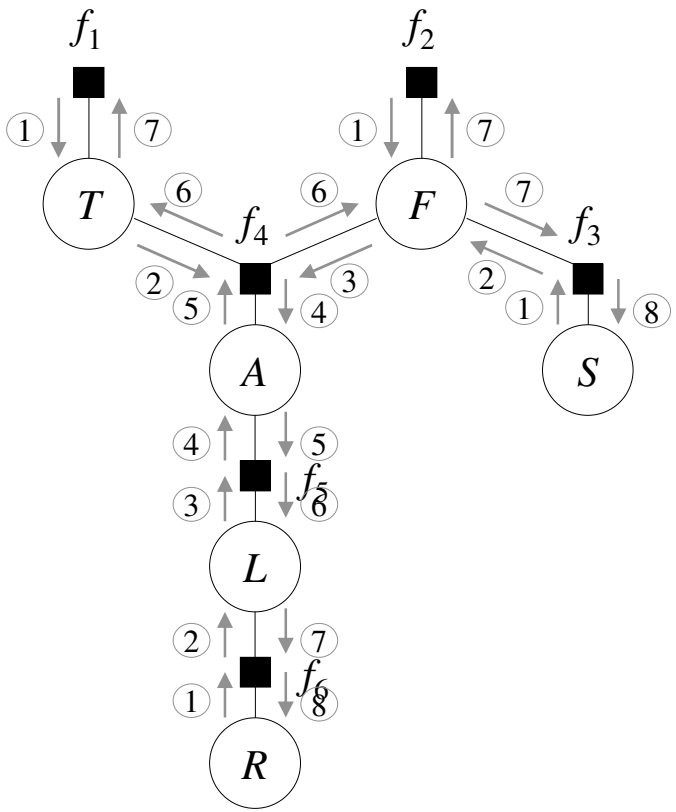
Da nodo variabile a nodo funzione:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)/\{f\}} \mu_{h\to x}(x)$$

Da nodo funzione a nodo variabile:

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

Factor graphs



Lo *scheduling* dei messaggi può essere diverso.
In questo caso: *flooding schedule*

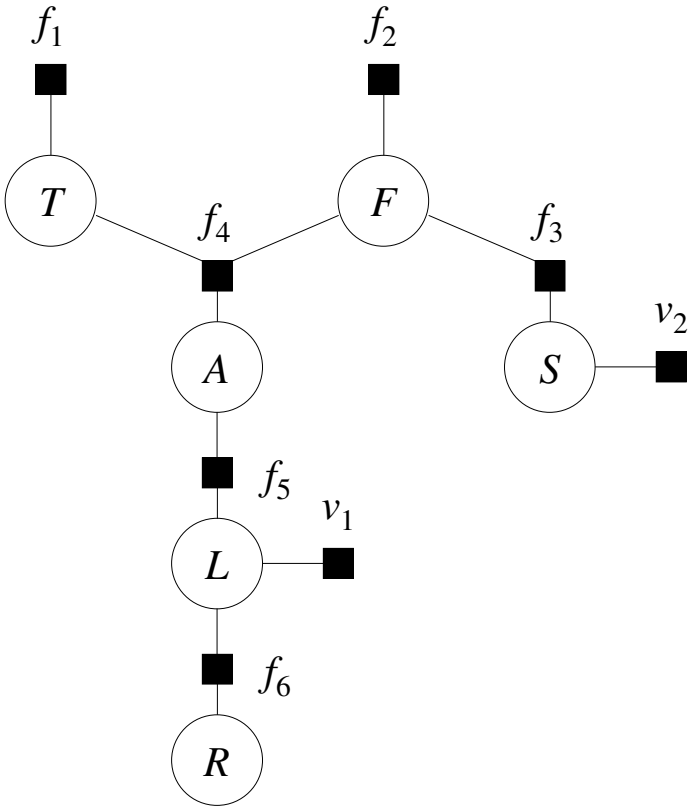
Message passing

In un grafo aciclico, ad un certo punto, lo scambio di messaggi si arresta

Le probabilità marginali associate a ciascun nodo variabile sono il prodotto dei messaggi ricevuti

$$P(x) = \prod_{f \in n(x)} \mu_{f \to x}(x)$$

Factor graphs



Message passing

In un grafo aciclico, ad un certo punto, lo scambio di messaggi si arresta

Le probabilità marginali associate a ciascun nodo variabile sono il prodotto dei messaggi ricevuti

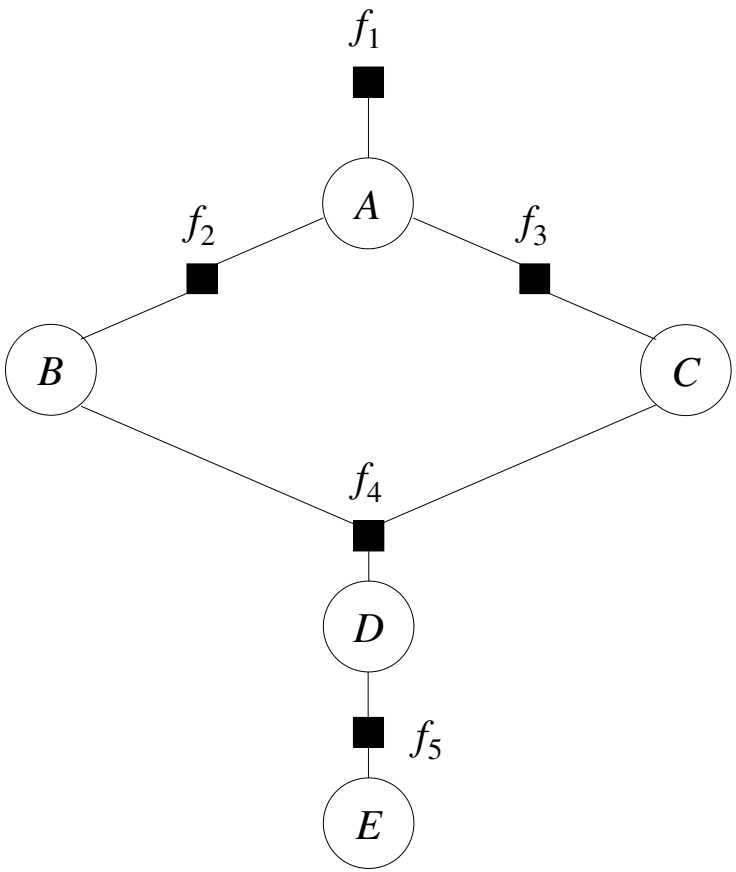
$$P(x) = \prod_{f \in n(x)} \mu_{f \to x}(x)$$

Le osservazioni si rappresentano introducendo nuovi nodi funzione, che esprimono un vincolo: $v(x) = 1$ per il valore osservato e 0 altrimenti

Il metodo deve essere nuovamente eseguito

I valori ottenuti per i nodi variabile devono essere normalizzati, per ottenere le probabilità marginali condizionali

Factor graphs con cicli



Message passing

In un grafo ciclico, in generale,
lo scambio di messaggi non si arresta mai

In alcuni casi non esiste nemmeno
un punto di partenza

Variante del metodo

- Inizialmente, ciascun nodo invia
  un messaggio con un vettore di valori 1
- L'invio dei messaggi è ripetuto, al cambiare
  di uno dei messaggi ricevuti

Non c'è garanzia di convergenza, il metodo
può continuare all'infinito

In molti casi pratici, tuttavia, ciò avviene:
dopo un certo tempo, lo scambio si stabilizza
e i messaggi non cambiano più