



**UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO**

DIPARTIMENTO DI INFORMATICA
CORSO DI LAUREA TRIENNALE IN INFORMATICA

Progetto ICON

MedicoOnline

<https://github.com/marck9918/MedicoOnline.git>

**Calcolo delle inferenze su una Kb
Malattie mediante
programmazione Logica ProbLog
e confronto con un modello
Bayesiano**

Marco Fusillo [698872]

Prof. Nicola Fanizzi

Scopo del progetto

Questo progetto nasce dall'esigenza di conoscere quale patologia si soffre dall'evidenza di alcuni sintomi. Lo scopo è quello di effettuare inferenza su una knowledge base che restituisca con quale probabilità si soffre di una possibile patologia. Con il presente lavoro non si vuole sostituire la figura professionale del medico, anzi, è fortemente raccomandato un suo consulto, questo progetto è solamente uno strumento diagnostico. Il seguente lavoro non presenta nessuna attendibilità dal punto di vista dati a causa dell'assenza di una figura che vi potesse fornire informazioni attendibili. I dati inseriti sono da osservare come da "segnaposto" per evidenziare il funzionamento del metodo proposto. Il progetto è stato implementato mediante programmazione logica problog e mediante reti bayesiane.

Realizzazione del progetto

Il progetto è stato realizzato interamente in python tranne per la scrittura della knowledge base che è stata scritta in problog mediante un editor online vedi <https://dtai.cs.kuleuven.be/problog/editor.html>. In python le librerie utilizzate sono le seguenti:

- Problog: Libreria per l'implementazione problog per l'installazione in ambiente windows: **`pip install problog`**
- Pgmpy: Libreria per l'implementazione della rete bayesiana per l'installazione in ambiente windows: **`pip install pgmpy`**
- Datetime: Libreria per estrapolare la data di sistema

Scelte Progettuali

L'implementazione della knowledge base si è pensato di implementarla mediante il linguaggio di programmazione logica probabilistico: problog. Il prolog è un linguaggio di programmazione basato su regole che permette di rappresentare e ragionare sulla conoscenza, ma se volessimo rappresentare l'incertezza questo non vi è possibile con la versione base. Qui entra in soccorso il problog che non è altro che un linguaggio basato su prolog. Problog estende la sintassi del prolog introducendo sostanzialmente fatti e clausole probabilistiche. Vedi esempio:

0.5 :: non_essere_promosso

Il seguente fatto scritto in problog indica che la probabilità dell'atomo *non_essere_promosso* sia 0.5. È possibile chiedere quale sia la prob di tale fatto mediante:

query(non_essere_promosso)

Inoltre, problog introduce anche due predicati fondamentali per effettuare inferenza ossia: **query** che serve per effettuare una ricerca probabilista del goal ed **evidence** che permette di asserire gli atomi osservati per il calcolo della probabilità condizionato dell'obbiettivo.

Per calcolare la probabilità di un goal si utilizzano due regole per propagare le probabilità definite nella base di conoscenza:

- La prob di una testa di una clausola è pari alla produttoria delle probabilità degli atomi del corpo (se non specificata)
Ad es.
febbre:- sudorazione, brividi
0.3 :: sudorazione
0.2 :: brividi
query(febbre) -> 0.6
Altrimenti la probabilità sarà quella specificata.
- Le probabilità di un atomo per ciascun mondo possibile vengono sommate per mondo possibile si intende l'istanziamento della kb ove ciascuna clausola a una e sola testa
Ad es.
febbre: - tosse.
febbre: - nausea.

0.1 :: tosse.
0.3:: nausea.

query(febbre). -> 0.037

Ogni clausola probabilistica scritta nella kb è nella seguente forma:

nomeMalattia :- sintomo₁ and sintomo₂ ... sintomo_n

Inoltre, per ogni sintomo è stata definita la propria probabilità nella seguente forma:

nomeSintomo :- evidence(stag (ST), true),

(

ST=estate, nomeSintomo (probabilità);

ST=autunno, nomeSintomo(probabilità);

ST=inverno, nomeSintomo (probabilità);

ST=primavera, nomeSintomo (probabilità)

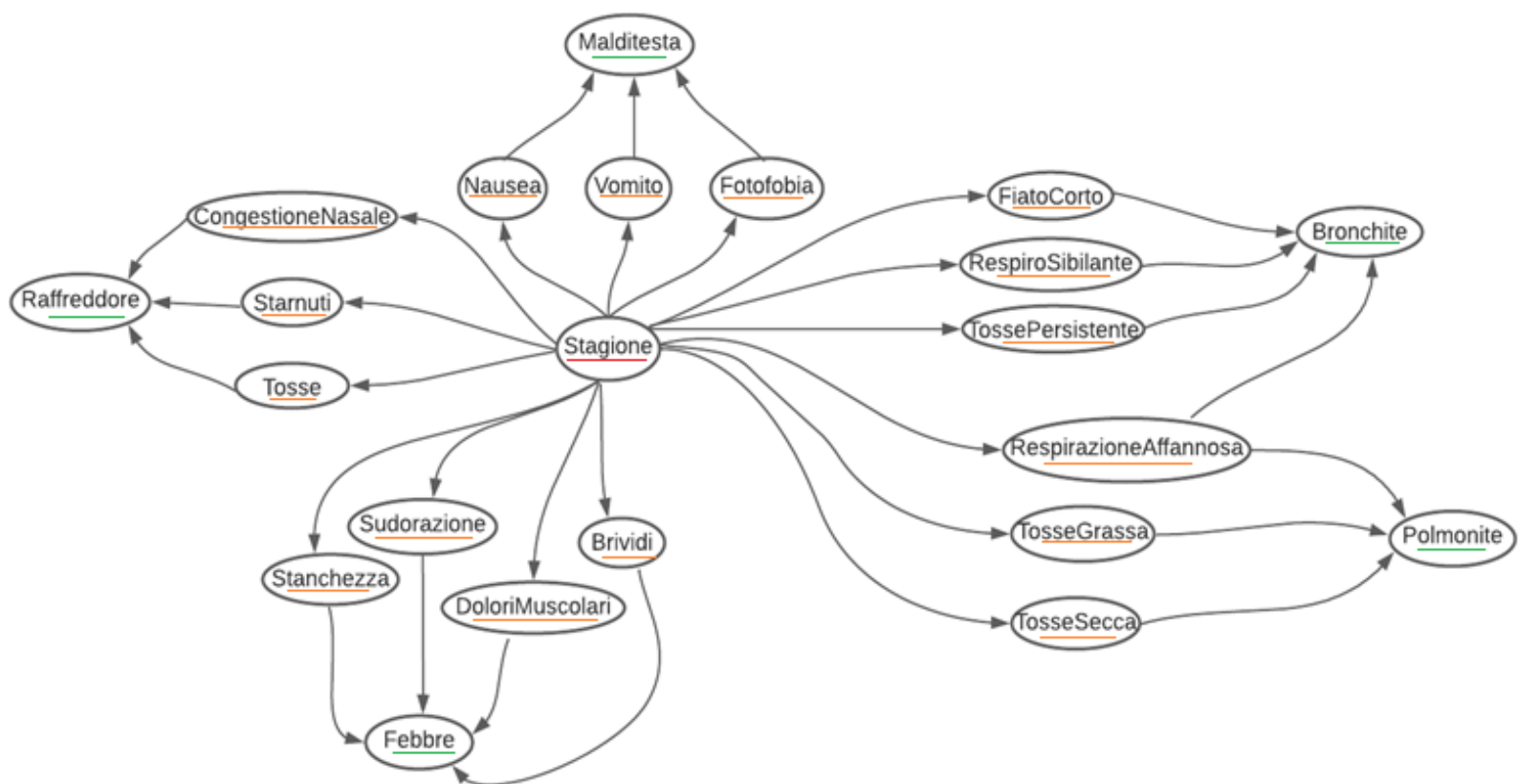
).

È stata effettuata tale scelta per differenziare le diverse probabilità di avere un sintomo in base alle diverse stagioni. Le diverse stagioni vanno ad influenzare sostanzialmente la probabilità di avere un sintomo X e quindi la probabilità di avere una malattia X. Ad es. è molto più probabile avere un raffreddore l'inverno invece che l'estate, quindi, è più probabile "soffrire" di congestione Nasale l'inverno invece che l'estate. La composizione della base di conoscenza permette un tipo di inferenza che va dagli effetti alle cause. Differentemente dall'altro modello proposto che è possibile un'inferenza in tutti e due i versi dalle osservazioni delle cause alla probabilità degli effetti.

Si è pensato di rappresentare lo stesso problema mediante un altro approccio probabilistico ovvero reti bayesiana. La rete bayesiana è un modello adeguato al lavoro in un dominio incerto. Non disponendo di un dataset la creazione della rete è stata effettuata manualmente seguendo le seguenti “regole”:

1. Si sceglie un ordinamento delle variabili X_1, \dots, X_n
2. Si scelgono le variabili genitori ossia quelle variabili che influenzano altre variabili del problema, basandosi su relazioni intuitive e sull’ordinamento
3. Costruzione delle cpt

Detto ciò, è possibile visionare graficamente la rete bayesiana realizzata mediante la Fig.1 ove si evincono le relazioni tra nodi.



Come è possibile evincere:

- Il nodo padre del modello Bayesiano: Stagione
- I nodi sottolineati di arancione i nodi genitori
- I nodi sottolineati di verde i nodi foglia

Inoltre, per ogni nodo vengono definite le CPT Conditional Probability Tables, ogni riga contiene per ogni valore del nodo la probabilità condizionata rispetto ad una determinata combinazione dei genitori. La somma delle righe delle probabilità condizionate deve essere pari a 1. Un nodo padre possiede solamente una riga ove vi è specificato il valore che la variabile possa assumere. Di seguito sono illustrate le cpt di ogni nodo:

P(stag)	
ESTATE	0.25
PRIMAVERA	0.25
AUTUNNO	0.25
INVERNO	0.25

	Estate		Autunno		Inverno		Primavera	
	T	F	T	F	T	F	T	F
tossePersistente	0.3	0.7	0.7	0.3	0.8	0.2	0.4	0.6
respiroSibilante	0.4	0.6	0.7	0.3	0.6	0.4	0.5	0.5
fiatoCorto	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
tosseGrassa	0.2	0.8	0.7	0.3	0.8	0.2	0.4	0.6
tosseSecca	0.3	0.7	0.8	0.2	0.5	0.5	0.5	0.5
fotofobia	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
vomito	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
nausea	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
stanchezza	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
doloriMuscolari	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
brividi	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
sudorazione	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
congestioneNasale	0.3	0.7	0.6	0.4	0.7	0.3	0.7	0.3
starnuti	0.1	0.9	0.5	0.5	0.6	0.4	0.5	0.5
tosse	0.3	0.7	0.7	0.3	0.9	0.1	0.6	0.4
respirazioneAffanosa	0.3	0.7	0.6	0.4	0.7	0.3	0.8	0.2

(fiatoCorto) and (respiroSibilante) and (tossePersistente) and (respirazioneAffannosa) = R				P(bronchite R)
T fiatoCorto	T tossePersistente	T respiroSibilante	T respirazioneAffanosa	0.98
T fiatoCorto	T tossePersistente	T respiroSibilante	F respirazioneAffanosa	0.47
T fiatoCorto	T tossePersistente	F respiroSibilante	T respirazioneAffanosa	0.49
T fiatoCorto	T tossePersistente	F respiroSibilante	F respirazioneAffanosa	0.26
T fiatoCorto	F tossePersistente	T respiroSibilante	T respirazioneAffanosa	0.25
T fiatoCorto	F tossePersistente	T respiroSibilante	F respirazioneAffanosa	0.19
T fiatoCorto	F tossePersistente	F respiroSibilante	T respirazioneAffanosa	0.24
T fiatoCorto	F tossePersistente	F respiroSibilante	F respirazioneAffanosa	0.15
F fiatoCorto	T tossePersistente	T respiroSibilante	T respirazioneAffanosa	0.64
F fiatoCorto	T tossePersistente	T respiroSibilante	F respirazioneAffanosa	0.24
F fiatoCorto	T tossePersistente	F respiroSibilante	T respirazioneAffanosa	0.18
F fiatoCorto	T tossePersistente	F respiroSibilante	F respirazioneAffanosa	0.17
F fiatoCorto	F tossePersistente	T respiroSibilante	T respirazioneAffanosa	0.24
F fiatoCorto	F tossePersistente	T respiroSibilante	F respirazioneAffanosa	0.2
F fiatoCorto	F tossePersistente	F respiroSibilante	T respirazioneAffanosa	0.14
F fiatoCorto	F tossePersistente	F respiroSibilante	F respirazioneAffanosa	0.09

(stanchezza) and (brividi) and (sudorazione) and (doloriMuscolari) = R				P (febbre R)
T stanchezza	T brividi	T sudorazione	T doloriMuscolari	0.89
T stanchezza	T brividi	T sudorazione	F doloriMuscolari	0.48
T stanchezza	T brividi	F sudorazione	T doloriMuscolari	0.45
T stanchezza	T brividi	F sudorazione	F doloriMuscolari	0.25
T stanchezza	F brividi	T sudorazione	T doloriMuscolari	0.4
T stanchezza	F brividi	T sudorazione	F doloriMuscolari	0.34
T stanchezza	F brividi	F sudorazione	T doloriMuscolari	0.16
T stanchezza	F brividi	F sudorazione	F doloriMuscolari	0.12
F stanchezza	T brividi	T sudorazione	T doloriMuscolari	0.4
F stanchezza	T brividi	T sudorazione	F doloriMuscolari	0.3
F stanchezza	T brividi	F sudorazione	T doloriMuscolari	0.24
F stanchezza	T brividi	F sudorazione	F doloriMuscolari	0.2
F stanchezza	F brividi	T sudorazione	T doloriMuscolari	0.25
F stanchezza	F brividi	T sudorazione	F doloriMuscolari	0.18
F stanchezza	F brividi	F sudorazione	T doloriMuscolari	0.15
F stanchezza	F brividi	F sudorazione	F doloriMuscolari	0.1

(tosseGrassa) and (tosseSecca) and (respirazioneAffannosa) = R			P(polmonite R)
T tosseGrassa	T tosseSecca	T respirazioneAffannosa	0.99
T tosseGrassa	T tosseSecca	F respirazioneAffannosa	0.4
T tosseGrassa	F tosseSecca	T respirazioneAffannosa	0.6
T tosseGrassa	F tosseSecca	F respirazioneAffannosa	0.24
F tosseGrassa	T tosseSecca	T respirazioneAffannosa	0.34
F tosseGrassa	T tosseSecca	F respirazioneAffannosa	0.2
F tosseGrassa	F tosseSecca	T respirazioneAffannosa	0.18
F tosseGrassa	F tosseSecca	F respirazioneAffannosa	0.1

(fotofobia) and (vomito) and (nausea) = R			P(maldisenteria R)
T Fotofobia	T Vomito	T Nasuea	0.9
T Fotofobia	T Vomito	F Nasuea	0.45
T Fotofobia	F Vomito	T Nasuea	0.57
T Fotofobia	F Vomito	F Nasuea	0.24
F Fotofobia	T Vomito	T Nasuea	0.6
F Fotofobia	T Vomito	F Nasuea	0.3
F Fotofobia	F Vomito	T Nasuea	0.3
F Fotofobia	F Vomito	F Nasuea	0.2

(tosse) and (starnuti) and (congestioneNasale) = R			P(raffreddore R)
T Tosse	T Starnuti	T CongestioneNasale	0.89
T Tosse	T Starnuti	F CongestioneNasale	0.44
T Tosse	F Starnuti	T CongestioneNasale	0.65
T Tosse	F Starnuti	F CongestioneNasale	0.2
F Tosse	T Starnuti	T CongestioneNasale	0.67
F Tosse	T Starnuti	F CongestioneNasale	0.36
F Tosse	F Starnuti	T CongestioneNasale	0.24
F Tosse	F Starnuti	F CongestioneNasale	0.1

Nelle reti bayesiane è possibile effettuare due tipologie di inferenza, stiamo parlando:

- Inferenza esatta
- Inferenza approssimata

L'inferenza esatta è possibile rappresentarla in pgmpy mediante la tecnica Variable Elimination: si salvano dei valori in variabile intermedie e ad ogni passo si riutilizzano le variabili utilizzate per facilitare la computazione. Oltre alla variabile Elimination è possibile applicare inferenza approssimata la quale può essere utile nel caso di una rete più complessa in un eventuale estensione di tale progetto. L'inferenza approssimata può essere di due tipi:

- Rejection Sampling
- Fordward Sampling

Nel caso di studio è stato applicata ApproxInference che non altro che una classe che applica inferenza approssimata attraverso fordward sampling nel caso di query senza evidence o attraverso rejection sampling altrimenti.

Sviluppi futuri

- Un possibile sviluppo futuro potrebbe essere quello di introdurre un servizio di allerta automatico nei confronti del medico competente o delle guardie mediche in caso il sistema predica determinate malattie ad alto rischio.
- Un'altra possibile estensione potrebbe essere quella di aiutare il “paziente” nella ricerca delle malattie, in particolare nella ricerca dei sintomi: ad es. il paziente soffre di febbre ma non riesce a descrivere i propri sintomi; quindi, il sistema pone al paziente una serie di domande tali che lo possano aiutare a capire i propri sintomi
- Un'altra possibile estensione sarebbe quella di introdurre un aggiornamento automatico del sistema con malattie sconosciute

Conclusioni

Nel seguente lavoro sono stati proposti due approcci probabilistici per l'implementazione di un sistema diagnostico *MedicoOnline*. Uno è basato su programmazione logica problog e l'altro sulle reti bayesiane. Naturalmente bisogna di includere nello sviluppo esperti del dominio, i quali saranno in grado di inserire con certezza la probabilità di avere una determinata malattia con determinati sintomi. Inoltre, vi si osserva che l'approccio mediante Reti bayesiane comporta una maggiore libertà nel calcolo delle inferenze, ma d'altronde richiede un numero di parametri maggiori rispetto al problog. Il primo approccio (Problog) risulta essere semplice ed intuitivo ma non permette flessibilità nel calcolo inferenziale.