Soft Computing Laboratory

Technical Report
TR-001-A

PROCESSI INFERENZIALI PROBABILISTICI
E SISTEMI ESPERTI

Domenico Calabrò e Giangiacomo Gerla

Febbraio 2002

Dipartimento di Matematica e Informatica
University of Salerno
84081 Baronissi – Salerno – ITALY

PROCESSI INFERENZIALI PROBABILISTICI E SISTEMI ESPERTI

Domenico Calabrò e Giangiacomo Gerla
Dipartimento di Matematica ed Informatica Università di Salerno
Via S. Allende Baronissi (SA)

Abstract. An framework for expert systems probabilistic in nature is proposed. Such a framework arises from an idea exposed in Gerla [1994]a which is related with the notions of Boolean valuation and similarity. A brief description of a prototype based on Visual Basic and Access languages is given.

Keywords: Sistemi esperti, probabilità, indiscernibilità.

1. Introduzione.

In Bacchus [1990] e in Halpern [1990] sono confrontate due affermazioni del tipo:
(i) " un uccello può volare " ;
(ii) " Tweety (un particolare uccello) può volare"
e pongono in rilievo la loro diversità quando si voglia assegnare ad esse delle valutazioni probabilistiche. Infatti l'asserzione che la probabilità di (i) è 0.9 esprime una informazione di tipo statistico circa il rapporto tra il numero di uccelli capaci di volare ed il numero totale degli uccelli. Più problematico invece è dire che cosa significa l'assegnare a (ii) la probabilità 0.9. Ciò perché le valutazioni probabilistiche sembrerebbero possibili solo in relazione a collettivi (punto di vista statistico) o successioni (frequentismo) e comunque non ad una affermazione che, come questa, si riferisce ad un singolo caso. D'altra parte è un fatto che due sole sono le possibilità:

- Tweety è capace di volare e quindi (ii) è vera
- oppure Tweety non è capace di volare e quindi (ii) è falsa.

Apparentemente più appropriato sembra il punto di vista soggettivista per il quale la probabilità è una misura di credenza e quindi l'assegnare una probabilità ad (ii) è una scelta soggettiva non formalizzabile in alcun modo. Tale punto di vista sarebbe confermato dal fatto che due individui diversi, con esperienze differenti e con informazioni differenti circa Tweety, potrebbero dare valutazioni diverse per (ii). Tuttavia un tale punto di vista rappresenta, più che una soluzione al problema, una rinuncia di fatto a capire quali siano i meccanismi consci o inconsci che possono portare una persona ad una data valutazione probabilistica. Conseguentemente, rappresenta anche una rinuncia a riprodurre tali meccanismi nella costruzione di un sistema esperto.
In Gerla [1994]a si affronta un tale problema riscrivendo (ii) come segue

"un uccello (scelto a caso) con le stesse caratteristiche osservabili di Tweety è capace di volare"

e quindi di assumere che l'attribuzione della probabilità 0.9 equivale e dire che:
"il 90% di tutti gli uccelli che soddisfano le stesse caratteristiche osservabili di Tweety sono capaci di volare".

Una tale proposta deriva dal convincimento che i nostri "gradi di credenza" nascano dall'esperienza circa una grande classe di casi (passati) che consideriamo simili a quello (attuale) esaminato: esperienza che è immagazzinata nella nostra mente anche se non necessariamente in maniera cosciente e quantizzata. Ciò giustifica anche la soggettività dei gradi di credenza e la relativa variabilità di tali gradi da individuo ad individuo. Infatti ciascuno può avere

- un accumulo diverso di esperienze (relative a casi passati),
- una maggiore o minore conoscenza delle caratteristiche del caso sotto considerazione,
- un differente criterio per stabilire quando considerare "simili" due casi, (criterio dipendente da quali caratteristiche si ritengono rilevanti nello stabilire che due casi sono simili).

In Gerla, Calabrò Scarpati, [2001] si tenta di tradurre considerazioni epistemologiche di tale tipo in uno strumento per sistemi esperti capaci di elaborare informazione di natura probabilistica. A tale scopo si pongono una serie di concetti utili per costruire un motore inferenziale.

In questo articolo si sviluppano ulteriormente tali idee. Inoltre, si descrive un prototipo di sistema esperto recentemente implementato nel laboratorio di Soft-Computing per testare le potenzialità applicative dei concetti proposti. Nella costruzione di tale prototipo abbiamo utilizzato Visual Basic ed Access. Comunque un qualunque linguaggio visuale collegato ad un data-base relazionale è altrettanto adatto allo scopo.

2. Incompletezza, analogia e valutazioni booleane.

Per esporre le linee generali di quanto intendiamo fare lasciamo la questione di Tweety ed analizziamo, ad esempio, il procedimento di stima della probabilità di morte di un individuo, il Sig. Rossi, in un contratto assicurativo. Sia *D* l'asserzione "il Sig. Rossi vivrà ancora 30 anni" e supponiamo di voler dare una valutazione probabilistica di *D*. Allora appare naturale considerare la percentuale dei "casi analoghi" al Sig. Rossi che hanno vissuto più di 30 anni e per fare questo è necessario coinvolgere due fonti di informazione.

Una fonte è relativa al Sig. Rossi (caso attuale); verranno verificate alcune sue caratteristiche che siano da un lato facilmente rilevabili e dall'altro significative per la determinazione della speranza di vita. Ad esempio l'età, il sesso, il tipo di lavoro ed altro; chiameremo "osservabili" tali proprietà. Naturalmente risulta non osservabile la proprietà *D* che più interessa per la determinazione del premio di assicurazione.

Un'altra fonte è di natura statistica e deriva dalla osservazione di un numero sufficientemente grande di "casi passati". Per ciascuno di tali casi sono state rilevate sia le proprietà osservabili che la proprietà *D*. Tale tipo di informazione rappresenta la conoscenza generale che si possiede, e quindi, in un certo senso, la teoria generale da applicare.

Detto ciò, una valutazione probabilistica di *D* potrebbe essere ottenuta al modo seguente:

1) si appurano le proprietà "osservabili" del Sig. Rossi,

2) si isolano nell'archivio dei dati statistici di cui si è in possesso la classe dei "casi analoghi" al Sig. Rossi, cioè la classe degli individui che verificano le stesse proprietà osservabili di Rossi (valutazione booleana)

3) si assume come probabilità di *D* il rapporto tra il numero di casi analoghi che verificano *D* diviso il numero totale di casi analoghi (valutazione numerica successiva alla valutazione booleana).

In questo modo di procedere emergono alcuni punti che appaiono fondamentali in ogni logica probabilistica.

*i) **Distinzione tra osservabili e non osservabili***.

Se si avesse accesso completo alle informazioni relative a Tweety non avrebbe senso parlare di probabilità di una proposizione come "Tweety è capace di volare". Infatti potremmo affermare subito se tale proposizione è vera oppure è falsa. La necessità di ricorrere ad una valutazione probabilistica scatta per il fatto che spesso si considerano proprietà non accessibili all'osservatore. L'espressione "non accessibile" deve essere intesa in modo ampio e si riferisce sia a proprietà non conoscibili nel mondo attuale per mancanza di strumenti adeguati (ad esempio la capacità di volare di Tweety), sia a proprietà che si realizzano nel futuro (come la morte entro 30 anni del Sig. Rossi nel caso delle assicurazioni). Nel seguito pertanto distingueremo proprietà osservabili e proprietà non osservabili. Le attribuzioni di probabilità sono valutazioni di proprietà non osservabili che si avvalgono delle informazioni date dalle proprietà osservabili.

ii) Ruolo dell'analogia.

Una misura di probabilità è sempre una misura di un collettivo, pertanto quando ci si riferisce ad un caso particolare bisogna collegare tale caso a quello di un opportuno collettivo. Tale collettivo è costituito dall'insieme dei casi che si considerano analoghi al caso considerato. In altre parole, le valutazioni probabilistiche (di proprietà non osservabili) concernenti "casi singoli attuali" come quello di Tweety non sono altro che valutazioni relative a collettivi di "casi analoghi al caso attuale" esaminati in passato. Pertanto il concetto di "analogia" gioca un ruolo centrale nella probabilità.

iii) Struttura booleana dell'informazione.

La valutazione numerica della probabilità di una proprietà non osservabile *D* è successiva alla valutazione booleana che associa a *D* l'insieme dei casi analoghi che verificano *D*. La necessità di coinvolgere le valutazioni booleane è un punto fondamentale poiché solo sui valori booleani è possibile, in un certo senso, elaborare calcoli. Si ricordi in proposito che la logica probabilistica, pur essendo simile alle logiche a più valori per il fatto di assegnare alle formule valutazioni numeriche diverse da quelle, classiche, di tipo 0-1, ha una caratteristica peculiare: non è una logica di tipo vero-funzionale. Ciò significa che il valore di verità di una formula composta non dipende funzionalmente dai valori di verità delle proposizioni componenti. Questa caratteristica di tipo semantico si trasmette naturalmente anche all'apparato sintattico creando insormontabili limiti alla possibilità di dedurre il valore numerico della probabilità di una asserzione quando si conoscano solo i valori numerici delle probabilità di altre asserzioni.

Al contrario, come osserveremo nel prossimo paragrafo, le valutazioni booleane sono di tipo vero-funzionale e quindi un loro coinvolgimento rende possibile l'elaborazione di un calcolo.

3. Logica probabilistica.

Ricordiamo alcune definizioni fondamentali della logica probabilistica. Nel seguito indicheremo con L un linguaggio del calcolo proposizionale e con F l'insieme delle relative formule.

Definizione 3.1. Chiamiamo *valutazione probabilistica* di L una qualunque funzione $p:F\to[0,1]$ tale che:
i) $p(\alpha)=0$ per ogni contraddizione α ;
ii) $p(\alpha)=1$ per ogni tautologia α ;
iii) $p(\alpha\vee\beta)=p(\alpha)+p(\beta)$ se $\alpha\wedge\beta$ è una contraddizione;
iv) $p(\alpha)=p(\beta)$ se α è logicamente equivalente a β.

Le valutazioni probabilistiche non sono vero-funzionali. Infatti non è difficile provare il seguente teorema.

Proposizione 3.2. Detta $p:F\to[0,1]$ una distribuzione di probabilità sull'insieme di formule, esiste una funzione f tale che $p(\alpha\wedge\beta)=f(p(\alpha),p(\beta))$ per ogni coppia di formule α e β solo se p assume valori solo nell'algebra di Boole $\{0,1\}$.

Dim. Se esistesse una funzione f tale che $p(\alpha\wedge\beta)=f(p(\alpha),p(\beta))$, allora f sarebbe crescente e tale che $f(x,x)=x$ per ogni $x\in[0,1]$. Pertanto da $y\geq x$ dovrebbe derivare $f(x,y)\geq f(x,x)=x$. Detta α una qualunque formula, sia $p(\alpha)=x$, $p(-\alpha)=1-x$, e supponiamo che $x\leq 1-x$. Allora, avremmo che $0=p(\alpha\wedge-\alpha)=f(x,1-x)\geq x$ e quindi $x=0$. Analogamente si dimostra che se $x\geq 1-x$ allora $x=1$.

A differenza dalle valutazioni probabilistiche, le valutazioni booleane sono vero-funzionali e quindi per definirle è sufficiente limitarsi alla valutazione delle variabili proposizionali.

Definizione 3.3. Data un'algebra di Boole ***B***, una *valutazione booleana* è una funzione $v:F\to\boldsymbol{B}$ dall'insieme delle formule F in ***B*** tale che:

$$v(\alpha\wedge\beta)=v(\alpha)\wedge v(\beta)\ ;\ v(\alpha\vee\beta)=v(\alpha)\vee v(\beta)\ ;\ v(-\alpha)=1-v(\alpha)\ .$$

Chiamiamo *valutazione **B**-probabilistica* una terna $(\boldsymbol{B},v,p)$ con v valutazione booleana e p probabilità finitamente additiva definita in ***B***.

La seguente proposizione pone in rilievo come le nozioni di valutazione probabilistica e di valutazione ***B***-probabilistica siano interscambiabili.

Proposizione 3.4. Sia $(\boldsymbol{B},v,p)$ una valutazione ***B***-probabilistica e definiamo p' ponendo $p'(\alpha)=p(v(\alpha))$ per ogni formula α. Allora p' è una valutazione probabilistica che chiamiamo *associata* a $(\boldsymbol{B},v,p)$. Inoltre ogni valutazione probabilistica si può ottenere in questo modo.

Dim. La prima parte della proposizione è di immediata verifica. Sia p' una valutazione probabilistica e denotiamo con ***B*** l'algebra di Lindenbaum associata al linguaggio L. Si definisca inoltre v come la funzione che ad ogni formula α associa la relativa classe completa di equivalenza $[\alpha]$ e si ponga $p([\alpha])=p'(\alpha)$. Allora è immediato che $(\boldsymbol{B},v,p)$ è una valutazione ***B***-probabilistica la cui valutazione probabilistica coincide con p'.

4. Sistemi di rappresentazione della conoscenza.

In questo paragrafo seguiremo un formalismo simile a quello proposto in Pawlak [1982], e ciò specialmente per quanto riguarda la nozione di indistinguibilità. Le due definizioni principali che proponiamo corrispondono a due fasi di raccolta delle conoscenze.

Informazione sui casi reali.
Supponiamo che in una prima fase si ci limita a raccogliere informazioni circa una serie di casi sui quali si ha disponibilità completa di informazione.

Definizione 4.1. Chiamiamo *sistema di rappresentazione della conoscenza* una 4-pla
S=(*CR*, *AT*, *OSS*, v) dove
- *CR* è un insieme i cui elementi chiamiamo *casi reali*,
- *AT* è un insieme i cui elementi vengono detti *attributi*,
- *OSS* è un sottoinsieme di *AT*, detto l'insieme degli attributi *osservabili*
- v:$CR \times AT \to \{0,1\}$ è una funzione, detta *funzione informazione*.

Il significato che attribuiamo alla funzione di informazione v è che se v(c,α)=1 allora la proprietà α è risultata vera per il caso c, altrimenti è risultata falsa. Per semplicità supporremo che tutti gli insiemi coinvolti siano finiti e non vuoti. Ad ogni insieme *A* di attributi viene associata una relazione di equivalenza sull'insieme dei casi al modo seguente.

Definizione 4.2. Sia *S* = (*CR*, *AT*, *OSS*, v) un sistema di rappresentazione della conoscenza e *A* un insieme di attributi, allora si definisce una relazione binaria $\cong_A$ in *CR* ponendo

$$a \cong_A b \text{ se e solo se } v(a,\alpha)=v(b,\alpha) \text{ per ogni } \alpha \in A.$$

Se a $\cong_A$ b diremo anche che a e b sono *A-indiscernibili*. Nel caso di *AT*-indiscernibilità parleremo semplicemente di indiscernibilità.

In altri termini due casi sono considerati *A*-indiscernibili se verificano le stesse proprietà in *A*. Ovviamente la relazione $\cong_A$ è di equivalenza e quindi possiamo associare ad ogni caso c la relativa classe di equivalenza $[c]_A$.

Compattazione delle informazioni.
Supponiamo che un sistema di rappresentazione della conoscenza sia stato ottenuto da una raccolta di informazioni relativa ad un insieme di casi passati. Il data-base, che dovrebbe raccogliere le informazioni volute, che chiameremo per intenderci iniziale, dovrebbe però essere enormemente grande per essere affidabile, tanto grande da non poter essere maneggiato. Possiamo però procedere ad uno suo snellimento, infatti, rispetto al processo inferenziale che ci interessa e che esporremo nel seguito, non è di nessuna utilità distinguere due casi che siano indiscernibili. Pertanto, appare naturale effettuare un quoziente del sistema iniziale rispetto alla relazione di indiscernibilità e sostituire ad S il relativo quoziente S*=(*CR**, *AT*, *OSS*, v*) dove $CR^*=\{[r]_A \mid r \in CR\}$ è il quoziente di *CR* modulo la relazione di indiscernibilità e v* è definita ponendo $v^*([c]_A,a)=v(c,a)$ per ogni $c \in CR^*$. Gli elementi di *CR** verranno detti *casi tipo* (mentre abbiamo chiamato casi reali gli elementi di *CR*).
Le dimensioni di tale sistema sarebbero abbastanza contenute; infatti poiché due casi tipo si devono sempre differenziare per almeno una proprietà, detto n=|*AT*| la cardinalità di *AT*, il numero di casi tipo è inferiore a 2^n. Se si procede in tale modo è però necessario, per non perdere informazioni essenziali, memorizzare per ogni caso-tipo t il numero w(t) di casi che esso rappresenta (il "peso" da dare al caso-tipo). Si perviene in tale modo alla seguente definizione.

Definizione 4.3. Chiamiamo *struttura inferenziale probabilistica*, in breve *sip*, una struttura del tipo S = (*CT*, *AT*, *OSS*, v, w) tale che
- (*CT*, *AT*, *OSS*, v) sia un sistema di rappresentazione della conoscenza,
- due elementi di *CT* sono sempre discernibili
- w:$CT \to N$ una funzione a valori interi detta *funzione peso*.

Come abbiamo già detto, chiameremo *caso-tipo* ogni elemento di *CT*. Inoltre chiameremo *peso totale* di una *sip* S il numero $w(S)=\sum\{w(c) \mid c \in CT\}$, cioè il numero dei casi reali rappresentati globalmente da S. Nel seguito indicheremo con L il linguaggio del calcolo proposizionale il cui insieme delle variabili è *AT* mentre indicheremo con L* il linguaggio il cui insieme di variabili è *OSS*. Naturalmente, dato un caso c, è possibile estendere la valutazione v(c,_) ad ogni formula del linguaggio L al solito modo, ponendo

$$p(c,\alpha\wedge\beta)=\min\{p(c,\alpha),p(c,\beta)\} \quad , \quad p(c,\alpha\vee\beta)=\max\{p(c,\alpha),p(c,\beta)\} \quad , \quad p(c,-\alpha)=1-p(c,\alpha).$$

Proposizione 4.4. Ogni *sip* determina una valutazione ***B***-probabilistica (***B***,vb,vp) di L tale che:
- ***B*** è l'algebra di Boole delle parti di *CT*:
- $vb(\alpha)=\{c \in CT \mid v(c,\alpha)=1\}$:
- vp:$\boldsymbol{B} \to [0,1]$ è la probabilità definita dall'uguaglianza vp({c})=w(c)/w(S).

In particolare ogni *sip* determina una valutazione probabilistica delle formule, che denoteremo ancora con vp, in cui, per ogni formula α,

$$\mathrm{vp}(\alpha) = \frac{\sum \{\mathrm{w(c)} / \mathrm{v(c}, \alpha) = 1\}}{\mathrm{w(S)}},$$

cioè vp(α) rappresenta la percentuale di casi reali in cui α risulta vera nella base di dati iniziale.

Definizione 4.5. Indichiamo con CT_{OSS} il quoziente di *CT* modulo la *OSS*-indistinguibilità e chiamiamo *modello non determinista* ogni elemento di tale insieme.

Naturalmente, dato un modello non determinista $[c]_{oss}$, ogni fomula α di L_{oss} sarà valutata in maniera classica (0 oppure 1). Invece se α è una formula in L che coinvolge alcuni non-osservabili, allora in generale la sua valutazione è di tipo probabilistico e quindi diversa da 0 ed 1. Infatti possono esistere alcuni casi-tipo in $[c]_{oss}$ che verificano α ed altri che non la verificano. Precisamente abbiamo che

$$\mathrm{vp}([c]_{oss}, \alpha) = \frac{\Sigma\{\mathrm{w(x)} \mid \mathrm{x} \in [\mathrm{c}]_{oss} \ \ \mathrm{v(x}, \alpha) = 1\}}{\Sigma\{\mathrm{w(x)} \mid \mathrm{x} \in [\mathrm{c}]_{oss}\}}.$$

cioè una valutazione probabilistica di α si ottiene considerando la percentuale dei casi-reali *OSS*-analoghi a c in cui α è verificata.

Nota: L'accumularsi dell'esperienza determina una variazione del *sip* di partenza. Infatti se si ha conoscenza di un nuovo caso c di cui si abbia conoscenza completa, allora dal *sip* iniziale si ottiene un nuovo *sip* nel seguente modo. Se c è discernibile dai casi già considerati allora si aggiunge c come nuovo caso-tipo con peso uno. Se c è indiscernibile ad uno dei casi tipo già presenti, allora si provvede solo ad aumentare di uno il peso di tale caso-tipo.

5. Il processo inferenziale.

Una volta che si è raccolto in un *sip* l'informazione "storica" sull'insieme dei casi passati, in presenza di un caso nuovo da esaminare si procede ad accumulare tutte le informazioni che si riescono ad avere su questo caso. Poiché le nuove informazioni che abbiamo assunto devono essere esprimibili nel linguaggio L_{oss} degli osservabili, è utile in proposito la seguente definizione.

Definizione 5.1. Chiameremo *teoria probabilistica* ogni teoria in L_{oss}, cioè ogni insieme *T* di formule nel linguaggio degli osservabili. Dato un *sip* S, porremo $T(CT)=\{c\in CT \mid v(c,\alpha)$ per ogni α in $T\}$ e diremo che *T* è *soddisfacibile* in S se *T*(*CT*) è diversa dal vuoto, cioè se esiste in S un caso-tipo soddisfacente *T*. *T* è *categorica* in S se tutti i casi-tipo verificanti *T* sono *OSS*-indiscernibili, cioè se *T* ammette un solo modello non determinista in S.

Le teorie categoriche corrispondono al possesso della massima informazione possibile relativamente al linguaggio degli osservabili.

Definizione 5.2. Sia S = (*CT*, *AT*, *OSS*, v, w) un *sip* e *T* una teoria probabilistica soddisfacibile in S, chiamiamo *sip definito da T* in S il *sip*

$$T(\mathrm{S})=(T(CT), AT, OSS, \mathrm{v}, \mathrm{w}).$$

In altre parole data una teoria probabilistica *T* ed un *sip* S, viene definito un *sip* *T*(S) che si ottiene da S eliminando tutti i casi-tipo che non verificano *T*. A sua volta tale *sip* determina una valutazione ***B***-probabilistica $(P(T(CT)),vb_T,vp_T)$ delle formule di L rispetto alla quale

- $vb_T(\alpha)=T(CT)\cap vb(\alpha)$ è l'insieme dei casi verificanti *T* intersecato con l'insieme dei casi verificanti α
- $vp_T(\alpha)$ è la percentuale dei casi passati verificanti α nell'ambito dei casi verificanti *T*.

In altre parole, $vp_T(\alpha)=vp(\alpha/T(CT))$, cioè $vp_T(\alpha)$ è la probabilità di α condizionata dall'evento *T*(*CT*):

$$\mathrm{vp}_T(\alpha) = \frac{\Sigma\{\mathrm{w(c)} \mid \mathrm{c} \in \mathrm{T(CT)},\ \mathrm{v(c},\ \alpha) = 1\}}{w(T(S))} = \mathrm{vp}(\alpha \ / \ T(S)).$$

Concludiamo questo paragrafo con la definizione principale. Sia S=(*CT*, *AT*, *OSS*, v, w) una struttura inferenziale probabilistica e sia ca un fissato modello di L_{oss} (cioè una valutazione del linguaggio L_{oss}) che chiameremo *caso attuale* (il Sig.Rossi del primo paragrafo). Chiamiamo *evidenza* di ca l'insieme *T*(ca) delle formule in L_{oss} vere in ca. Supponiamo inoltre che *T*(ca) sia soddisfacibile in S, cioè che in S vi sia almeno un caso-tipo *OSS*-equivalente a ca.

Definizione 5.3. Se α è una formula in L chiameremo probabilità che ca verifichi α la probabilità di α nel *sip* associato a *T*(ca).

Pertanto, coerentemente a quanto accenato nella introduzione:
la probabilità che un caso singolo ca verifichi una proprietà α è uguale alla percentuale dei casi OSS-analoghi a ca che in passato sono risultati verificare α.

6. Processo inferenziale step-by-step e strategie.
Abbiamo visto che per giungere ad una buona valutazione di una asserzione-diagnosi α si deve pervenire alla conoscenza dell'evidenza *T*(ca). Ciò si può ottenere facendo contemporaneamente tutte le domande relative ai predicati *OSS* ed in tale caso il processo inferenziale di un sistema esperto equivarrebbe al fare riempire un modulo con le informazioni del caso da esaminare ed alla successiva consultazione di una tabella statistica. Un tale modo di procedere potrebbe essere valido nell'esempio della determinazione della probabilità di morte del Sig. Rossi. In realtà nella maggior parte delle situazioni il comportamento di un esperto deve essere molto più raffinato e le domande devono essere fatte scegliendo una opportuna "strategia". L'evidenza *T*(ca) viene allora acquisita per passi successivi (e solo parzialmente) come frutto dell'accumulo di informazioni ottenute attraverso una opportuna successione di domande $Q_1,...,Q_k$ nel linguaggio L_{oss}. Indichiamo con R_i la formula Q_i se $Q_i \in T(ca)$ e $-Q_i$ se $Q_i \notin T(ca)$, e chiamiamo R_i la *risposta alla domanda* Q_i. Allora viene definita una successione di teorie probabilistiche $T_o,...,T_k$ tali che:
- $T_o=\phi$;
- $T_1=\{R_1\}$
. . .
- $T_{j+1}=T_j\cup\{R_{j+1}\}$.

Ancora, a tale successione di teorie corrisponde la successione di *sip*
S_o = S,
$S_1=T_1(S)$,
. . .
$S_k=T_k(S)$

Ricordiamo che tale successione si ottiene cancellando man mano da *CT* i casi-tipo non compatibili con le informazioni

$$CT_0 = CT,\ CT_1=T_1(CT),\ldots,\ CT_k=T_k(CT).$$

Nel *sip* finale S_k si possono leggere le probabilità di tutte le formule del linguaggio L stante l'informazione T_k ottenuta sul caso ca. Si noti che il passaggio da una teoria T_i alla sua estensione T_{i+1} tramite R_{i+1} corrisponde al passaggio dalla distribuzione di probabilità associata a T_i a quella condizionata dall'evento $vb(R_{i+1})$. In ogni passo del processo abbiamo delle valutazioni probabilistiche di tutte le formule, al crescere dell'informazione tali valutazioni divengo man mano più attendibili fino ad arrivare alla teoria finale la quale risulta categorica ed individua un modello probabilistico simile a quello attuale.

La strategia da adottare, cioè il criterio con cui tali domande devono essere man mano scelte, deve rispondere a vari scopi. Ovviamente, uno di questi è quello di pervenire al massimo di informazione con il numero minimo di domande. Ma vi sono altri aspetti a cui stare attenti. Se ad esempio ci riferiamo al problema di effettuare una diagnosi in ambito medico allora bisogna considerare prima le informazioni che si possono ottenere direttamente dal paziente sul momento e poi quelle che si traducono in analisi da fare. Inoltre nel proporre analisi risulta essenziale tenere conto dei costi, del fatto che possano danneggiare il paziente, del fatto che siano più o meno dolorose, dell'urgenza del caso, della rilevanza di una analisi per escludere o verificare una particolare diagnosi e così via. Naturalmente la strategia sarà interattiva, cioè il tipo di analisi proposto in una data fase del processo di diagnosi dipende dalle risposte alla domande-analisi precedentemente fatte. In definitiva la questione si pone nei seguenti termini che sono classici di una situazione di problem-solving :
data un sip S che sia espressione delle informazioni T possedute ad un dato stato del processo, quale delle possibili domande-mosse-del-gioco conviene tentare?

Purtroppo i fattori di cui si dovrebbe tenere conto per pesare una possibile scelta sono molti (ma forse non maggiori di quelli relativi ad una mossa di scacchi). E' probabile che un sistema esperto non si possa sostituire ad un medico nella valutazione complessiva di tali fattori e quindi nella scelta. Forse è però possibile chiedere ad un sistema esperto di quantificare in modo preciso alcuni di questi fattori.
Esamineremo solo il problema della riduzione al minimo del numero di domande. Una buona strategia deve tendere a giungere il massimo possibile di informazione utile sul caso in considerazione con il numero minore di domande. La ricerca di una tale strategia quindi coincide con quella relativa al famoso gioco di "indovinare un numero". Un giocatore ***B*** deve indovinare un numero n scelto a caso in un insieme finito I={x_1,...,x_n} e questo facendo il minor numero possibile di domande ad un giocatore A. Per comodità identifichiamo ogni domanda con un sottoinsieme Q di I. Si suppone inoltre che la scelta iniziale di n avvenga in accordo con una distribuzione di probabilità iniziale p e che quindi l'informazione iniziale sia data dall'entropia H(p) di p

$$H(p) = -\sum_{k=1}^{n} p(x_k)\cdot \ln p(x_k)\,,$$

con la solita convenzione per cui 0 · ln(0)=0. Si tratta di scegliere una domanda Q in modo che l'entropia diminuisca quanto più possibile in seguito alla risposta. Poiché è possibile avere sia una risposta positiva che una negativa, dobbiamo scegliere Q in modo che si abbia una diminuzione media dell'entropia.
Nel seguito indicheremo con H(Q) l'entropia che si avrebbe in caso di risposta positiva alla domanda Q, cioè l'entropia della probabilità p(_,Q) condizionata da Q e con H(-Q) l'entropia in caso di risposta negativa. Inoltre indicheremo con HM(Q) il valore di aspettazione dell'entropia in relazione alla domanda Q, cioè

HM(Q)=p(Q)H(Q)+p(-Q)H(-Q).

Infine chiamiamo *grado di rilevanza* di Q la quantità H(Q)= -p(Q)ln(p(Q))-p(-Q)ln(p(-Q).

Proposizione 6.1. Per ogni domanda Q risulta che

HM(Q)=H -H(Q).

Dim. Abbiamo che

$$H(Q) = -\sum_{k=1}^{n} p(x_k|Q)\cdot \ln p(x_k|Q)\,,$$

pertanto, tenendo conto del fatto che P(x/Q)=0 se x non appartiene a Q e p(x/Q)=p(x)/p(Q) altrimenti,

$$P(Q)H(Q) = -\sum_{x\in Q}^{n} p(x)\cdot(\ln p(x) - \ln p(Q)) = -\sum_{x\in Q}^{n} p(x)\cdot \ln p(x) + p(Q)\ln p(Q)$$

Sommando all'analoga quantità p(-Q)H(-Q) si ottiene

$$HM(Q) = -\sum p(x)\cdot \ln p(x) - (-p(Q)\ln p(Q) - p(-Q)\ln p(-Q))\,.$$

Corollario 6.2. Il valore di aspettazione dell'entropia assume valore minimo quando il grado di rilevanza H(Q) di Q assume valore massimo. A sua volta ciò avviene quando Q è *bilanciata* cioè quando p(Q) è il più vicino possibile a p(-Q) e quindi a 0.5.

In definitiva una buona strategia per indovinare il numero n consiste nella seguente procedura (non deterministica)
- si pone p_0=p e si sceglie Q_1 bilanciata rispetto a p_0
- si pone R_1=Q_1 se la risposta a Q_1 è positiva e R_1=-Q_1 altrimenti
- si pone p_1=p_0(_|R_1) e si sceglie Q_2 bilanciata rispetto a p_1

. . .

- si pone R_i=Q_i se la risposta a Q_i è positiva e R_i=-Q_i altrimenti
- si pone si pone p_i=p_0(_|R_i) e si sceglie Q_{i+1} bilanciata rispetto a p_i

. . .

Il processo termina quando si perviene ad una entropia nulla cioè ad una probabilità p_k che assume valore 1 in n e valore zero negli altri numeri. Si noti che p_i=p(_ | $R_1\cap...\cap R_i$) e che quindi la scelta di Q_{i+1} deve essere tale che p($R_1\cap...\cap R_i\cap Q$) deve essere il più vicino possibile a p($R_1\cap...\cap R_i$)/2. Se facciamo l'ipotesi per cui ogni volta l'insieme $R_1\cap...\cap R_i$ dei numeri che rimangono possibili possa essere diviso in due parti equiprobabili, allora ad ogni domanda l'entropia diminuisce di ln(2). Scelta come base di ln il numero 2, abbiamo allora che l'entropia diminuisce di una unità.

In definitiva, il valore successivo della parte intera dell'entropia iniziale rappresenta il numero di domande che in media debbono essere fatte per indovinare il numero n. Ad esempio nel caso $n=2^k$ e supposto che tutti i numeri sono equiprobabili, risulta che k coincide con l'entropia iniziale e quindi saranno necessarie k domande per indovinare il numero n. La strategia da adottare consisterà nel dividere in due parti equipotenti l'insieme dei numeri che, in considerazione delle risposte già ricevute, rimangono possibili.

Tornando al nostro processo inferenziale, al passo i-esimo siamo in presenza di un *sip* S_i con relativa distribuzione di probabilità su CT_i data dall'uguaglianza $vp_i(\{c\})=w(c)/w(S_i)$, per c in CT_i. L'entropia di tale probabilità rappresenta il grado di informazione posseduto al passo i-esimo. Dopo aver fatto le domande $Q_1,...,Q_i$, la domanda Q da scegliere deve essere bilanciata rispetto a vp_i e quindi tale che

$$\sum_{c \in vb_i(Q)} w(c)$$

sia il più possibile prossimo al valore $w(S_i)/2$.

7. Un prototipo.

Per testare le idee di questo lavoro è stato costruito un prototipo di sistema esperto che permette di effettuare valutazioni di non osservabili in base a successive domande su eventi osservabili. Il programma è scritto in Visual Basic ed accede ad un data base relazionale in Access. Il sistema fornisce previsioni nel senso che, ad ogni passo del processo inferenziale, viene data una distribuzione di probabilità sulle possibili risposte (riguardo alle proprietà non osservabili). Il database di partenza (casi reali) dovrà essere costituito da una tabella con *items* (persone, pazienti, auto, abitazioni, ecc.) che rispondano a certe caratteristiche che supporremo booleane.

Successivamente, il processo di raggruppamento di tutti i casi in classi di casi simili porterà ad individuare una nuova tabella con una probabilità di uscita per tutti i casi tipo. Tale tabella si evolverà mano mano che vengono introdotte nuove informazioni in una tabella sempre più piccola. Ad ogni nuova informazione verranno eliminati i casi non compatibili con essa.

Ad esempio: dato un insiemi di pazienti con vari sintomi, dobbiamo individuare se un nuovo paziente con una certa serie di sintomi ha una certa malattia o meno utilizzando la base di dati precedente.

Nel caso di un auto si deve poter determinare se sia conveniente o no assicurarla in base a particolari caratteristiche.

La base di dati dovrà essere costituita da una tabella contenente dei record che individuano ognuno un caso e che ha come campi oltre all'individuazione (nome paziente, nome malattia) i sintomi riscontrati.

Un'altra tabella dovrà contenere le domande e le risposte relative che per noi hanno un significato e la relativa posizione (come campo) nella tabella dei casi.

Avremo bisogno di due *form* per l'inserimento dei dati: una per le domande da eseguire (domande con risposta si o no) ed una per l'inserimento dei dati effettivi.

Avremo poi bisogno di una *form* che suggerisca le domande da eseguire (in base alla stategia di richiesta) e che determini in passi successivi i record che rispondano a quella domanda. Più precisamente, nel prototipo la strategia si basa solo sull'entropia e quindi sulla scelta di domande "bilanciate". Uno dei campi dovrà essere il numero dei casi reali che un dato caso tipo rappresenta ed un altro la probabilità di uscita di quel caso. Se il numero delle risposte possibili è n ogni caso avrà una n-pla di valori {0,1} corrispondenti alle risposte sulle domande possibili. Il numero dei casi possibili in base al numero di risposte n è rappresentato da uno spazio di n-ple costituito al massimo da 2^n elementi.

Ogni caso rappresenterà quindi uno o più casi possibili e tale numero verrà detto *rappresentazione*. In base alle rappresentazioni è possibile determinare la probabilità di appartenenza del caso in oggetto ad uno dei casi inseriti nel database. Effettuando ad ogni passo la somma delle probabilità dei casi rimanenti che hanno come risposta "si", sarà possibile determinare quella domanda che più si avvicina ad una probabilità 0.5.

Una volta eseguita la domanda e ridotto il numero dei casi possibili, nella tabella risultante bisognerà ricalcolare le probabilità in base ai casi rimasti, e quindi rideterminare la domanda successiva in base alla precedente strategia.

Strategia di individuazione della richiesta

Dato un insieme di casi Ci con determinate caratteristiche Ri possiamo per esempio individuare le seguenti tabelle.

Casi	R1	R2	R3	R4	Casi simili	Attesa
C1	1	0	0	0	3	3/20
C2	0	1	1	0	2	2/20
C3	1	1	1	0	7	7/20
C4	1	1	0	1	8	8/20
	18/20	17/20	9/20	8/20	20	

La domanda da effettuare da effettuare ad ogni passo dovrà essere quella che riduce al minimo l'entropia e che dimezza (più o meno) i casi possibili (probabilità 0.5) in modo che la risposta si o no ad una certa domanda suddivida in due l'insieme dei casi.

Si dimostra che tale strategia individua la risposta nel peggiore dei casi in un numero minore di tentativi.

In questo caso la domanda da effettuare è quella relativa al caso R3 perché è quella che più si avvicina a 0.5. Infatti divide in due il numero dei casi possibili.

Se la risposta è si (1) allora la tabella risultante è

Casi	R1	R2	X	R4	Casi simili	Attesa
C1	1	0	0	0	3	3/11
C4	1	1	0	1	8	8/11
	1	8/11	0	8/11		

In questo caso la domanda da fare è da scegliere tra la R2 e la R4 la cui risposta finale individuerà certamente un caso finale.

Nel caso in cui abbiamo una mancanza di informazioni il nostro sistema può continuare a funzionare?

Una ipotesi di lavoro in tal senso può essere quella di isolare innanzitutto tutti i casi che hanno mancanza di informazione, e presi singolarmente aggiungerli al sistema in modo da rispettare le probabilità di uscita dei casi simili.

In altre parole se abbiamo un record relativo ad un caso in cui manca qualche valore dobbiamo innanzitutto determinare le percentuali dei casi simili, ottenuti dando al valore indeterminato tutti i valori possibili. A questo punto il numero dei casi simili verrà incrementato non più di un'unità, ma semplicemente della relativa percentuale di casi simili già considerati.

Per esempio, se dovessimo inserire il seguente caso, in cui R2 è indeterminata:

Casi	R1	R2	X	R4
X	1	n	0	1

e considerando i due casi simili in cui R2 può assumere sia valore 0 che valore 1:

Casi	R1	R2	X	R4	Casi simili	Attesa
C1	1	0	0	0	2	1/5
C4	1	1	0	1	8	4/5
					10	

avremo che l'aggiunta del caso x all'insieme di tali casi produrrà un insieme di questo tipo:

Casi	R1	R2	X	R4	Casi simili	Attesa
C1	1	0	0	0	2.2	2.2/11
C4	1	1	0	1	8.8	8.8/11
					11	

Quindi fondamentalmente avremo aggiunto al sistema un solo caso, ripartendolo però nelle sue percentuali ai relativi casi simili.

Nel caso in cui gli elementi siano più di uno si ragiona nello stesso modo ed ottenendo un numero di casi simili pari a 2n dove n è il numero delle indeterminazioni.

Per esempio nel caso di due indeterminazioni avremo:

Casi	R1	R2	X	R4
x	1	n	0	n

considerando i quattro casi simili in cui R2 ed R4 possono assumere sia valore 0 che valore 1:

Casi	R1	R2	X	R4	Casi simili	Attesa
C1	1	0	0	0	20	1/5
C2	1	0	0	1	10	1/10
C3	1	1	0	0	30	3/10
C4	1	1	0	1	40	2/5
					100	

Avremo, quindi, che l'aggiunta del caso x all'insieme di tali casi produrrà un insieme di questo tipo:

	R1	R2	X	R4	Casi simili	Attesa
C1	1	0	0	0	20.2	202/1010
C2	1	0	0	1	10.1	101/1010
C3	1	1	0	0	30.3	303/1010
C4	1	1	0	1	40.4	404/1010
					101	

In tal modo il sistema potrebbe automaticamente gestire una indeterminazione di valori senza alterare fondamentalmente il sistema creato.

8. Alcuni particolari

Il prototipo di programma che permette di studiare il caso in esame è costituito da un programma scritto in Visual Basic, che accede ad un database Access contenente le tabelle casi, domande e risultati.

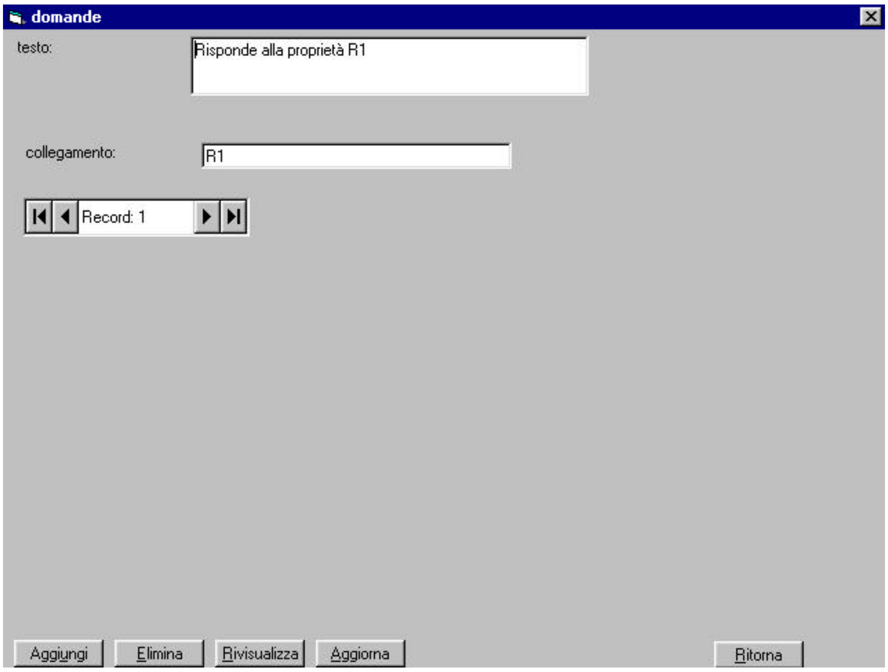
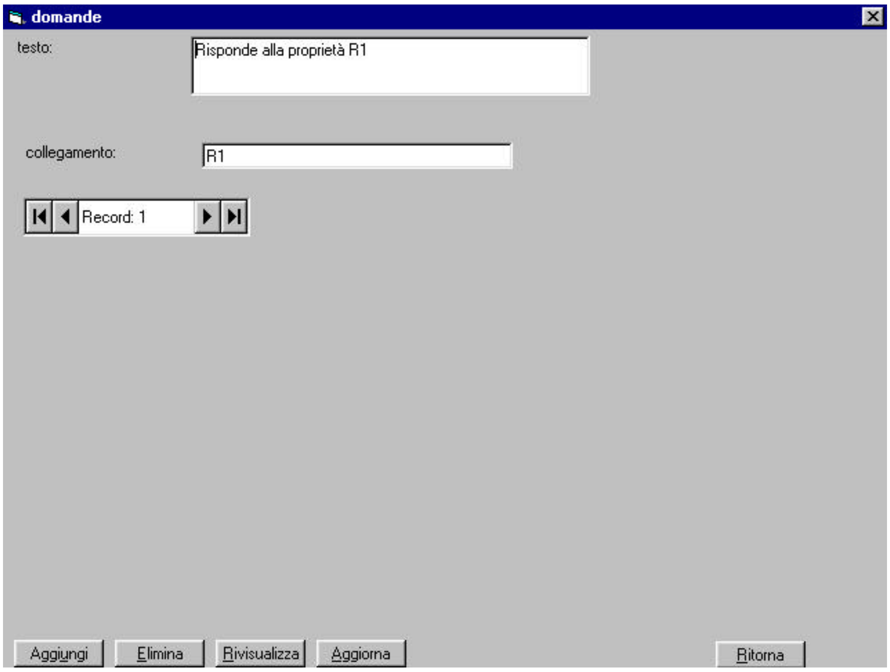


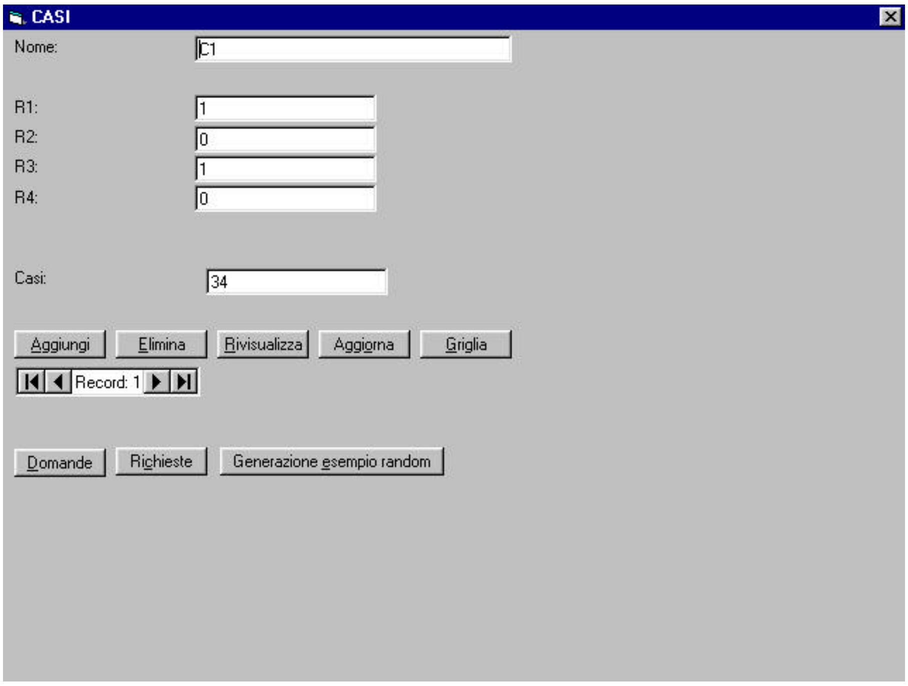
Tabella domande

Le domande da porre sui singoli casi saranno contenute nella tabella domande ed ognuna di esse sarà associata ad una risposta, in questo caso R1, R2, R3 o R4.



I casi

I casi trattabili vengono descritti nella tabella casi contenenti le risposte alle domande precedentemente inserite, in questo caso i casi saranno 16 poiché le risposte alle domande sono 4 e di tipo booleano.

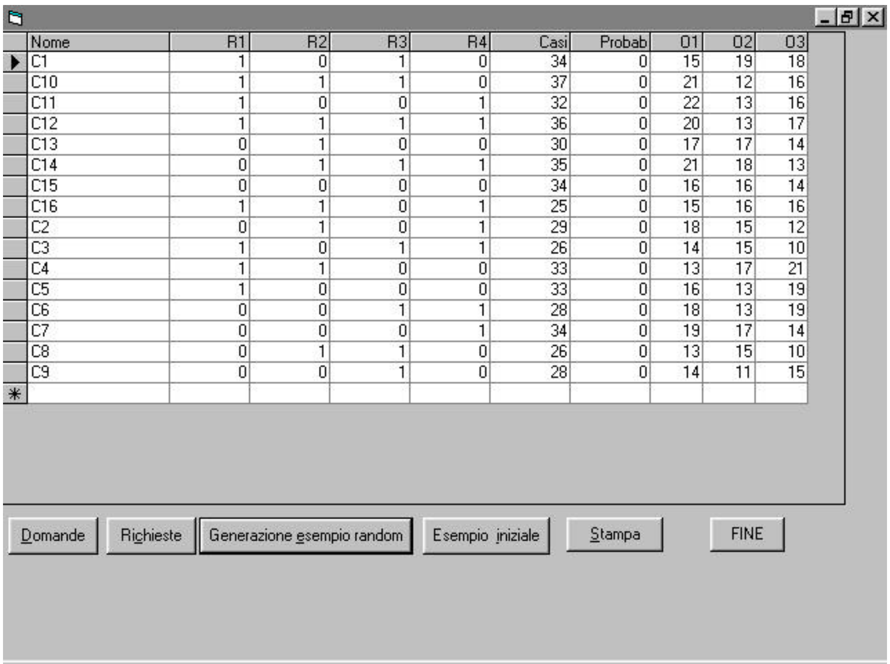


Generazione di casi random

Nella generazione random dei casi viene richiesto il numero dei casi da generare e successivamente viene creata una tabella contenente una serie di risposte booleane a tutte le riposte possibili, in questo caso ad R1, R2, R3 e R4 ed inoltre vengono inserite le colonne O1, O2 ed O3 relative agli obiettivi.

Successivamente tali casi generati vengono raggruppati i casi simili in base alle risposte possibili.

In questo caso essendo quattro i valori booleani possibili si ottengono 16 casi da cui ottenere un raffinamento fino ad ottenere il caso più probabile rispondente alla nostra soluzione.

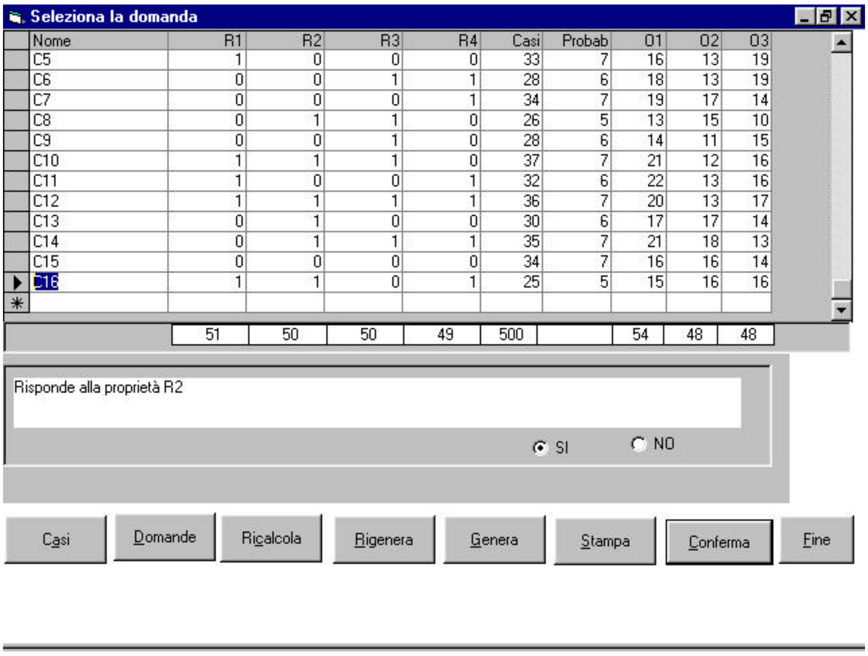


Nome	R1	R2	R3	R4	Casi	Probab	O1	O2	O3
C1	1	0	1	0	34	0	15	19	18
C10	1	1	1	0	37	0	21	12	16
C11	1	0	0	1	32	0	22	13	16
C12	1	1	1	1	36	0	20	13	17
C13	0	1	0	0	30	0	17	17	14
C14	0	1	1	1	35	0	21	18	13
C15	0	0	0	0	34	0	16	16	14
C16	1	1	0	1	25	0	15	16	16
C2	0	1	0	1	29	0	18	15	12
C3	1	0	1	1	26	0	14	15	10
C4	1	1	0	0	33	0	13	17	21
C5	1	0	0	0	33	0	16	13	19
C6	0	0	1	1	28	0	18	13	19
C7	0	0	0	1	34	0	19	17	14
C8	0	1	1	0	26	0	13	15	10
C9	0	0	1	0	28	0	14	11	15

Raffinamento

Eseguito il raggruppamento si procede al raffinamento premendo il bottone richieste a cui è associata una maschera dove verranno poste domande in successione, la cui conferma permetterà la riduzione dei casi fino ad ottenere un caso possibile che rappresenterà la soluzione del nostro problema.

La scelta della domanda viene eseguita in modo da ridurre al massimo l'entropia e quindi ridurre drasticamente il numero delle domande da porre per ottenere gli obiettivi.

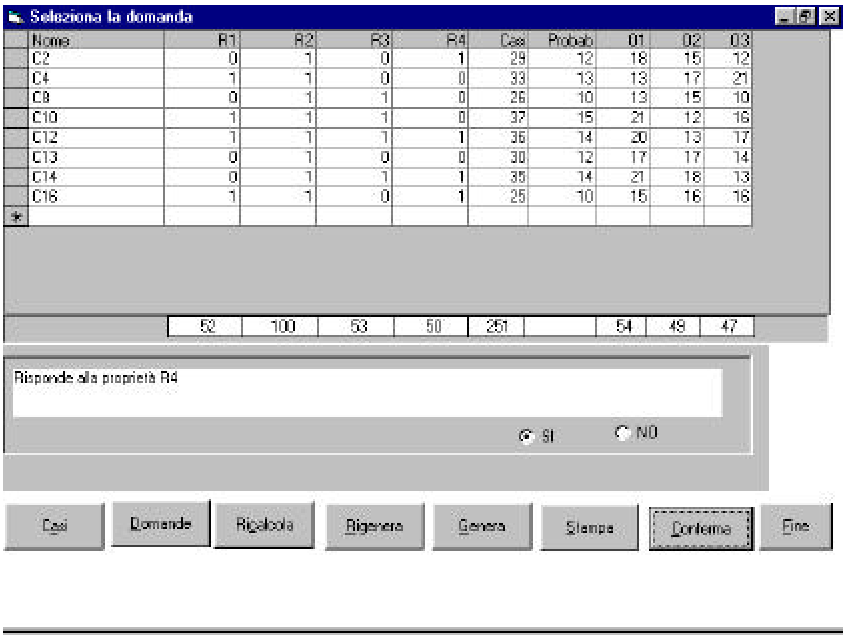


Nome	R1	R2	R3	R4	Casi	Probab	01	02	03
C5	1	0	0	0	33	7	16	13	19
C6	0	0	1	1	28	6	18	13	19
C7	0	0	0	1	34	7	19	17	14
C8	0	1	1	0	26	5	13	15	10
C9	0	0	1	0	28	6	14	11	15
C10	1	1	1	0	37	7	21	12	16
C11	1	0	0	1	32	6	22	13	16
C12	1	1	1	1	36	7	20	13	17
C13	0	1	0	0	30	6	17	17	14
C14	0	1	1	1	35	7	21	18	13
C15	0	0	0	0	34	7	16	16	14
C16	1	1	0	1	25	5	15	16	16
	51	50	50	49	500		54	48	48

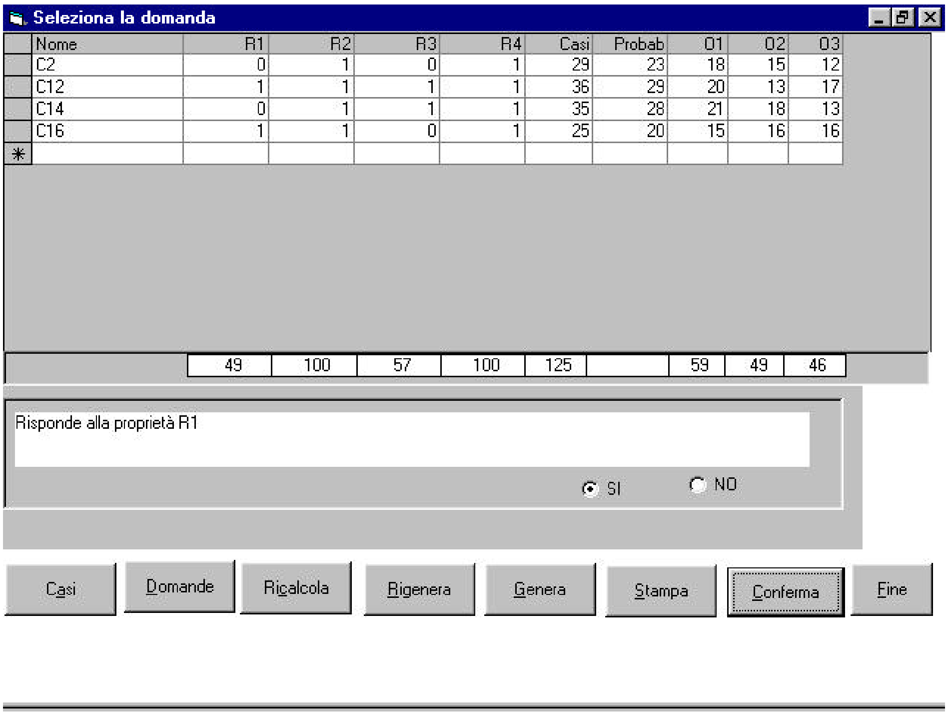
Riduzione dei casi

Come si può notare la risposta **si** alla domanda R2 riduce ad 8 i casi da considerare per la prossima domanda.

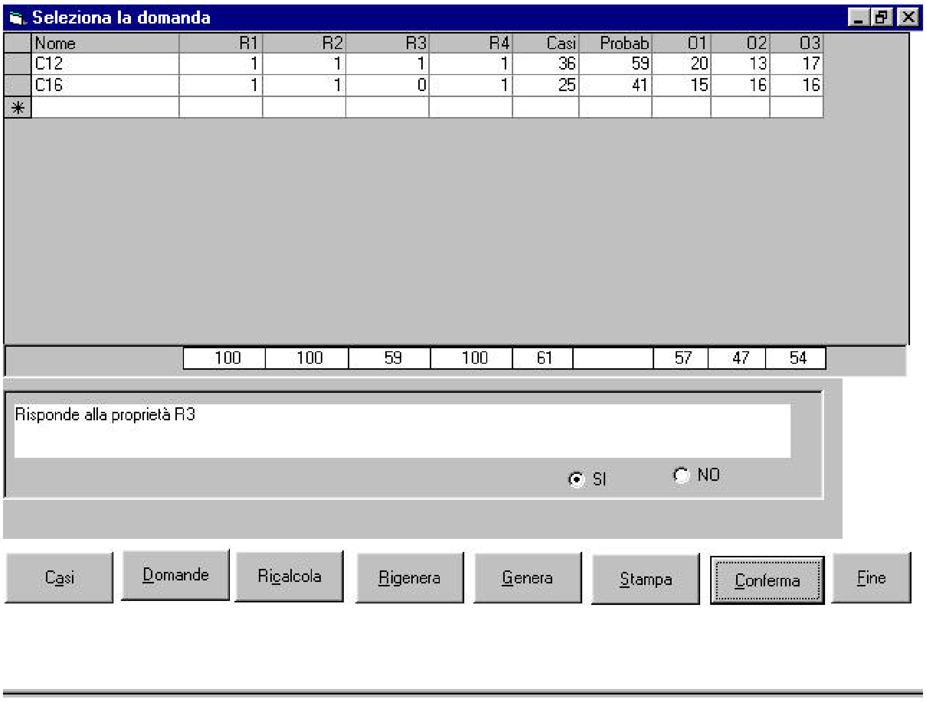
La risposta **si** riduce i casi a 4. Da notare che un'eventuale risposta negativa permetterebbe di ottenere sempre la riduzione a quattro casi ma ovviamente si otterrebbero come soluzione i quattro casi scartati precedentemente.



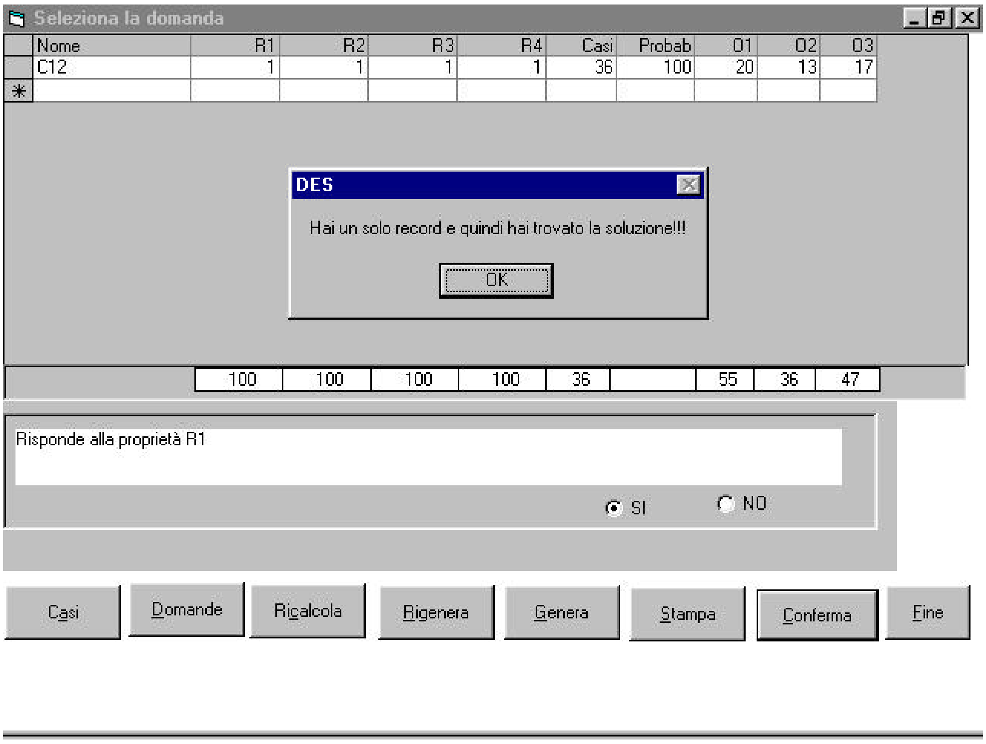
Nome	R1	R2	R3	R4	Casi	Probab	01	02	03
C2	0	1	0	1	29	12	18	15	12
C4	1	1	0	0	33	13	13	17	21
C8	0	1	1	0	26	10	13	15	10
C10	1	1	1	0	37	15	21	12	16
C12	1	1	1	1	36	14	20	13	17
C13	0	1	0	0	30	12	17	17	14
C14	0	1	1	1	35	14	21	18	13
C16	1	1	0	1	25	10	15	16	16
	52	100	53	50	251		54	49	47



Rispondendo ancora **si** a quest’ulteriore domanda si ottengono due soli casi.



La soluzione
Infine, rispondendo affermativamente a quest'ultima domanda si individua un unico caso che rappresenta la soluzione del problema.



Tale individuazione permette di giungere ad una conclusione relativa alle probabilità di raggiungimento degli obiettivi O1, O2 ed O3. Ciò significa che la probabilità maggiore permetterà di effettuare una previsione sul raggiungimento dell'obiettivo più probabile.

Bibliografia.

- F. Bacchus, [1990], Lp, a logic for representing and reasoning with statistical knowledge, *Comput. Intell.*, 6, 209-231.
- E. F. Codd, [1972], Further normalization in the database relational model, In R. Rushin ed. *Database Systems*, Princeton Hsall.
- G. Gerla, [1994]a, The probability that Tweety is able to fly, *International Journal of Intelligent Systems*, 9, 403-409.
- G. Gerla, [1994]b, Inferences in probability logic, *Artificial Intelligence* 70, 33-52.
- G. Gerla, [1994]c, Fuzzy refutations for probability and multivalued logics, *International J. of Approximate Reasoning* 4 (1994) 257-279.
G. Gerla, [1997], Probability-like functionals and fuzzy logic, *J. of Math. Anal. Appl.*, 216, 438-465.
- G. Gerla, [2000], *Fuzzy Logic: Mathematical Tools for approximate reasoning*, Kluwer Academic Press.
- G. Gerla, D. Calabrò, L. Scarpati [2001], Extension Principle and Probabilistic Inferential Process, in *Advances in Soft Computing*, Physica- Verlag.
- J.Y.Halpern, [1990], An analysis of first-order logic of probability, *Artificial Intelligence*, 46 (1990) 331-350.
- Z. Pawlak, [1982], Rough sets, *International Journal of Information and Computer Science*, 11, 344-356.