Web Mining: Analisi del Comportamento dell'Utente Internet Mediante Agenti Adattativi

Stefano Hajek
hajek@kame.usr.dsi.unimi.it

***Abstract* - I sistemi induttivi - sistemi che apprendono le proprie regole di funzionamento attraverso l'interazione con l'ambiente - si rivelano risorsa essenziale nella rappresentazione di dominii complessi e mutevoli.**

La necessità di reperire criteri guida attraverso i quali trattare l'enorme quantità di dati presenti in internet e la disponibilità di grande potenza di calcolo a basso costo stanno determinando la maturazione ed il passaggio di questi sistemi dall'ambito della sperimentazione a quello dell'applicazione reale: le soluzioni basate sul data mining estraggono e sintetizzano conoscenza da una massa di informazioni caotica e ridondante rappresentando in questo modo un concreto supporto alla decisione strategica.

Più in particolare, il marketing delle aziende che svolgono attività di vendita e promozione on line avanzano una forte richiesta di strumenti per l'analisi e l'interpretazione del comportamento dell'utenza web al fine di diversificare l'offerta di contenuti e prodotti ed adeguarla ad una domanda sempre più mobile e dispersa.

Questo articolo descrive un algoritmo basato sul paradigma della Genetic Programming che classifica degli indicatori statistici relativi all'attività degli utenti di un sito al fine di individuarne inclinazioni ed interessi: l'obbiettivo del presente lavoro è la ricerca di un metodo di selezione dei visitatori potenzialmente più assidui, sui quali cioè indirizzare un'azione di coinvolgimento mirato; il risultato del processo di analisi consisterà in una lista gerarchica di attributi che definiscano "l'utente fedele". La prestazione dell'algoritmo sarà valutata in termini di accuratezza della previsione e parsimonia della soluzione.

I. Il problema

Un importante sito web di una società del settore risparmio raccoglie dati relativi al comportamento di navigazione dei propri visitatori; tali dati vengono dapprima collezionati nel log file prodotto dal web server e successivamente aggregati in descrittori statistici relativi all'attività di ogni singolo utente.

N. medio sessioni/giorno	Numero di sessioni che l'utente compie mediamente nell'arco di una giornata
Durata media sessione	Durata media (espressa in minuti) di ogni sessione effettuata dall'utente
N. medio pagine visitate/sessione	Numero medio di pagine che l'utente visita durante una sessione
Tempo medio speso su ogni pagina	Tempo medio (espresso in secondi) che l'utente spende su ogni pagina
Fascia oraria 1-5	Principale fascia oraria di connessione dell'utente
Giorno della settimana	Giorni della settimana in cui l'utente compie principalmente le connessioni
Sequenza 1-10	Sequenza di pagine più visitata dall'utente tra le dieci più frequentate dall'intera utenza

Gli utenti registrati sono suddivisi in due classi: utenti sporadici (utenti che hanno fatto registrare fino a 15 visite dalla loro prima connessione) ed utenti assidui (utenti che hanno effettuato più di 15 visite); gli utenti che non risultano connessi almeno una volta nell'ultimo mese vengono rimossi dal campione.
Scopo dell'analista marketing è in questo caso l'identificazione di specifici valori dei descrittori statistici che distinguano i visitatori sporadici da quelli assidui; disponendo di tale discriminante è possibile vagliare nuovi utenti connessi intervenendo in anticipo sui fattori che ne incentivano l'interesse e dunque la fedeltà

Il modo più immediato per stimare la rilevanza di un particolare valore di un singolo parametro nel separare gli acquirenti dai non acquirenti è di contare l'occorrenza degli acquirenti che presentano quel valore per quel parametro: per un campione sufficientemente ampio tale conteggio rapportato agli utenti totali che soddisfano il medesimo requisito approssima la probabilità di incontrare un acquirente tra tutti gli utenti che dichiarano lo stesso valore.
Purtroppo però la valutazione di un singolo parametro raramente è significativa dal momento che differenti fattori intervengono simultaneamente nel determinare la propensione all'investimento originando una complessa interazione; ad esempio potremmo constatare che l'utenza pomeridiana è poco fedele, che chi frequenta la sequenza sette non è particolarmente costante ma che l'intersezione "utenti pomeridiani della sequenza sette" può rivelare un segmento interessante per la promozione di specifici programmi di "affiliazione" .

La Soft Computing offre diversi strumenti di classificazione multidimensionale: le reti neurali feed-forwarded sono buoni algoritmi di apprendimento supervisionato, ma, com'è noto, sono delle black-box ossia non esplicitano in termini leggibili per un esperto umano i modelli automaticamente generati.
Una classe di algoritmi di apprendimento con impostazione radicalmente differente ma che produce i propri risultati in forma di regole inferenziali in cascata è quella dei decision trees.
Gli algoritmi di questa famiglia suddividono gli individui in classi omogenee rispetto ad un valore target: essi esaminano attributo per attributo, valore per valore, e scelgono quello che consente la miglior separazione degli individui rispetto alla grandezza target associata; sui sottoinsiemi così ottenuti viene nuovamente applicata la medesima procedura,

iterandola sino ad ottenimento del desiderato livello di correttezza nella classificazione.
In particolare la misura di tale correttezza, ossia dell'omogeneità delle classi originate, è fornita dal valore di "entropia" (Quinlan 1993) dato da:

$$E = -P(A)\log_2[P(A)] - P(\neg A)\log_2[P(\neg A)] \quad (1)$$

Dove:

P(A) = probabilità che il valore target A risulti associato all'individuo di una classe
P(¬A) = probabilità che il valore target A non risulti associato all'individuo di una classe

Maggiore è l'entropia di una classe, maggiore è il suo "disordine" e minore è pertanto la correttezza della classificazione ottenuta.
In linea di principio l'algoritmo di generazione dei decision trees può essere semplicemente enunciato come segue:

1) suddividere gli individui di un campione in base al valore dell'attributo che procura il minimo valore di entropia rispetto ad un valore target associato;
2) ripetere dal passo 1) sui sottoinsiemi così ottenuti sino a riduzione del valore d'entropia ad una soglia fissata.

Il limite dei decision trees è determinato dalla procedura di suddivisione in classi per passi successivi: nulla impedisce infatti che la riduzione dell'entropia a livello locale - di singolo attributo - possa pregiudicare l'entropia complessiva del sistema (Miller 1990).
Per assicurare un'ottimizzazione globale del decision tree se ne può basare l'euristica di partizione su un sistema ad agenti adattativi: una popolazione di agenti, ciascuno dei quali rappresenta un modello discriminante, viene fatta pertanto competere sino a far emergere l'individuo che ottiene la maggior correttezza classificatoria.

II. Agenti intelligenti e Web Mining

Esiste una vasta letteratura relativa all'utilizzo di agenti nel web mining che tuttavia riguarda nel complesso un aspetto affine ma complementare a quello qui trattato, la tematica del data retrieval, che ben si presta ad essere affrontata tramite l'utilizzo di agenti intelligenti in grado di esplorare ed organizzare automaticamente i contenuti presenti in rete; ciònondimeno le linee di sviluppo tracciate in questo campo intercettano una serie di elementi base utili anche alla progettazione di un sistema per l'analisi della navigazione.
Tali linee di sviluppo possono essere così sintetizzate.

A. Intelligent Search Agents:

sviluppati per la ricerca e l'interpretazione di informazioni in domini specifici sfruttano una base di conoscenza predeterminata ed ottimizzata per quei particolari domini o per domini di argomento analogo (Brown et Al. 1994, Doorenbos et Al. 1996, Spertus 1997) .

B. Information Filtering / Categorization

Lo scopo di questi agenti è la classificazione per aree semantiche di differenti documenti per lo più mediante tecniche di clustering; da ciascun documento vengono estratte le parole chiave ritenute più rappresentative dello stesso sulla base di un criterio di ricorrenza, i diversi documenti vengono poi aggregati a seconda della corrispondenza tra le rispettive parole chiave (Frakes et Al. 1992, Chang et Al. 1997).

C. Personalized Web Agents

L'assegnazione di un contenuto ad un utente sulla base della probabilità che ne incontri l'interesse è il compito di questo tipo di agenti; anche in questo caso, come nel precedente i documenti vengono classificati per contenuto, tuttavia qui la classificazione risulta mirata sulle preferenze espresse dall'utente.
Questa famiglia di agenti è quella che presenta le maggiori analogie con quella da noi proposta, vuoi per l'utilizzo di un metodo di apprendimento supervisionato, vuoi per l'obbiettivo comune di indirizzare un messaggio ad un referente potenzialmente ricettivo (Oostendorp et Al. 1994, Armstrong et Al. 1995, Balabanovic et Al. 1995, Pazzani et Al. 1996).

Se nei lavori testè citati il profilo dell'utente coinvolto risulta in'ultima istanza definito dalle caratteristiche dei documenti prevalentemente consultati, nel presente lavoro il profilo dell'utente assiduo è definito dalle sue abitudini di navigazione; eccoci dunque approdati all'ambito di web mining che intendiamo qui focalizzare.

III. Agenti adattativi per la Web Analysis

Attenendoci all'accezione ampia ma rigorosa accolta da Holland (Holland 1995) faremo riferimento agli agenti adattativi come ad un complesso di unità in grado di far emergere dalla loro reciproca interazione un comportamento organizzato volto al raggiungimento di un obbiettivo comune; la misura di prossimità della colonia al conseguimento di tale obbiettivo, fornisce per retroazione un criterio di idoneità del comportamento colettivo e dunque di regolamentazione dello stesso.
Visto da una prospettiva più consona al nostro lavoro, un sistema di agenti può svolgere compiti di ricerca della strategia ottimale di soluzione di problemi complessi quale quello di classificare multidimensionalmente degli individui.

A. Rule system

Nel sistema qui proposto l'evoluzione dei modelli di classificazione avviene mediante un meccanismo di competizione / cooperazione tra agenti ciascuno dei quali è definito da una regola di trasformazione di un input (nel nostro caso differenti valori di differenti attributi) in un output (una stima del valore di assiduità).
Ciascuna regola di trasformazione risulta composta:
a) da una serie di nodi rappresentanti specifici valori di specifici attributi;
b) da dei rami rappresentanti le diverse alternative relative al soddisfacimento delle condizioni espresse da un valore di un attributo;
c) da delle foglie rappresentanti le classi della variabile obiettivo, classi mutualmente esclusive.

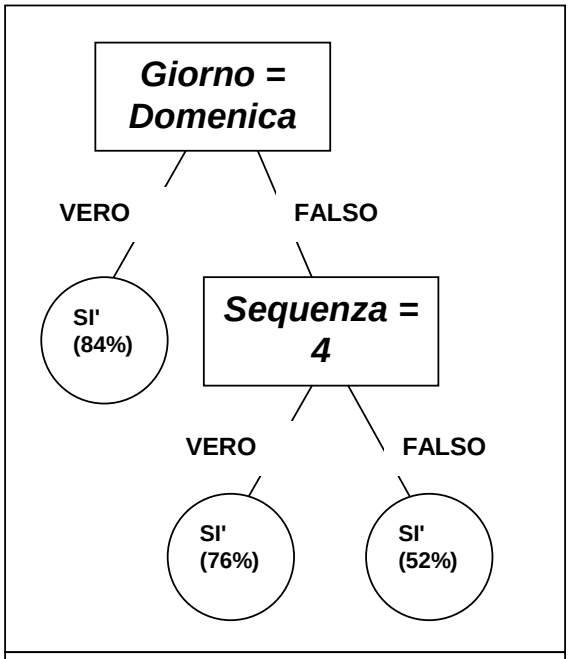


Figure 1 - Rappresentazione "ad albero" di una possibile soluzione rappresentata da un agente: ciascun nodo può rappresentare o il valore di un attributo discriminante, o il valore della variabile target associata; i rami rappresentano le alternative dell'esito

Dal punto di vista implementativo questo tipo di rappresentazione può essere ottenuta attraverso espressioni definite su un alfabeto di simboli di funzione, variabile, costante:

expr = if (x_i *R* c_i, t_1, t_2)

con

if = operatore condizionale
x_i *= simbolo di variabile*
c_i *= simbolo di costante*
R = {>,<,=}
t = {expr, 0, 1}

B. Generazione della popolazione iniziale

Gli individui (e dunque le regole) iniziali vengono prodotti per accrescimento iterativo: ciascun nodo generato casualmente può risultare o un attributo o una classe della variabile obiettivo; se è un attributo ammette una verifica ed origina pertanto tanti rami quanti sono le alternative della verifica, se invece è una classe non ammette alcun ramo ed arresta la crescita in quel punto.
E' evidente come da tale procedura derivino solamente espressioni sintatticamente corrette.

C. Credit Assignment

La competizione tra i diversi agenti viene attivata:
a) assegnando a ciascuno di essi un "valore adattativo" che ne misura l'efficacia rispetto al compito prefissato;
b) replicando gli individui più efficaci; al fine di enfatizzare il contributo degli individui valutati più promettenti viene infatti adottata la "steady state reproduction" la quale sostituisce gli individui peggiori In percentuale configurabile) con individui selezionati in quantità proporzionale al loro valore adattativo.
Il metodo più immediato per valutare in fase di addestramento il valore adattativo di un agente è quello che calcola il rapporto tra l'intero numero di classificazioni corrette ottenute e il numero totale di classificazioni (Vere 1995):

$$C = \frac{\sum_i n_i}{N} \quad (2)$$

dove: *C = correttezza complessiva; N = numero totale di individui classificati e* n_i *= numero di classificazioni corrette in i.*
Tuttavia questo canone di misurazione rischia di compromettere il delicato equilibrio tra complessità sintattica della soluzione e accuratezza della previsione: esso infatti non penalizza soluzioni con errori molto contenuti ma molto sofisticate che rischiano di rivelarsi troppo specifiche rispetto al contesto di apprendimento ed inadatte al contesto di verifica (non generalizzabili).
Per tale ragione in fase di generazione viene limitata la profondità degli alberi forzando l'algoritmo a cercare la miglior soluzione utilizzando un numero di ramificazioni contenute.
Un secondo intervento di incentivazione della compattezza dell'albero senza eliminazione di rami significativi è basato sulla conoscenza dell'utilità di ciascun sottoalbero la cui misura è data dalla misura dell'errore generato accoppiata alla misura dell'errore campionario; l'errore campionario dipende dalle dimensioni del campione scelto per indurre l'inferenza statistica: più popolato è il campione, più questo è rappresentativo dell'intero dominio.
Fissato un intervallo di confidenza l'errore atteso di generalizzazione viene calcolato come segue:

$$e_l = \frac{F_l + \frac{z^2}{2d_l} + z\sqrt{\frac{F_l}{d_l} - \frac{F_l^2}{d_l} + \frac{z^2}{4d_l^2}}}{\left(1 + \frac{z^2}{d_l}\right)} \quad (3)$$

Con

e_l *= errore di generalizzazione atteso alla foglia l*
d_l*= numero di campioni totali classificati nella foglia l*
F_l*= errore effettivo alla foglia l*
Z = numero di deviazioni standard corrispondenti ad un intervallo di confidenza fissato

Avendo in questo modo definito e_l possiamo sostituire n_l in (2):

$$n_l = (1 - e_l)d_l \quad (4)$$

D. Rule discovery

Alla fase della competizione selettiva segue quella di interazione cooperativa durante la quale la "conoscenza" (le regole) viene condivisa dagli agenti superstiti i quali vengono estratti casualmente a coppie e si scambiano porzioni di regole dando vita a nuove ipotesi, nuovi agenti

da avviare alla competizione; l'operatore di scambio (cross-over) preserva il requisito della chiusura, la morfologia dell'albero assicura infatti che la struttura risultante dall'innesto mantenga i criteri sintattici dell'albero originale.
Il crossover è il principale punto di forza del processo evolutivo: essa orienta implicitamente la ricerca poiché favorisce la propagazione di quei "blocchi" che, essendo appartenuti ad una buona soluzione, si possono supporre artefici del valore di quella soluzione; nel momento in cui viene valutato il materiale di una soluzione risultano implicitamente valutate tutte le soluzioni aventi qualche elemento in comune con essa (esplorazione parallela).
Al fine di evitare un'eventuale prematura convergenza della popolazione verso una soluzione non ottimale (un fenomeno di "conformismo sociale") l'operatore detto di "mutazione" percorre ricorsivamente l'albero alterando ogni nodo incontrato con una probabilità data da $P_m = x/lunghezza(albero)$, dove x è un parametro definito dall'utente; pertanto quando un nodo viene visitato esso può essere sostituito o da un nodo-attributo estratto casualmente o da un nodo-classe anch'esso estratto casualmente.

IV. RISULTATI

È stata verificata la qualità di GDT nel classificare i comportamenti di navigazione degli utenti web e nel prevedere il loro grado di assiduità a partire da alcuni indicatori statistici relativi all'attività di ciascun visitatore; la tabella che segue mostra i possibili valori dei diversi indicatori:

ATTRIBUTO	**VALORI**	**TIPO**
N. medio sessioni/giorno:	1 – 50	continua
Durata media sessione:	1 – 1200 (sec)	continua
N. medio pagine visitate/sessione:	1 – 50	continua
Tempo medio su ogni pagina:	1 – 1200 (sec)	continua
Fascia oraria:	1 – 5	discreta
Giorno della settimana:	1 – 7	discreta
Sequenza:	1 – 10	discreta

Ogni utente risulta quindi descritto dalla stringa dei valori assunti dai diversi attributi e da un valore target che è "1" se il visitatore si dimostra assiduo (più di quindici visite complessivamente effettuate), "0" altrimenti.

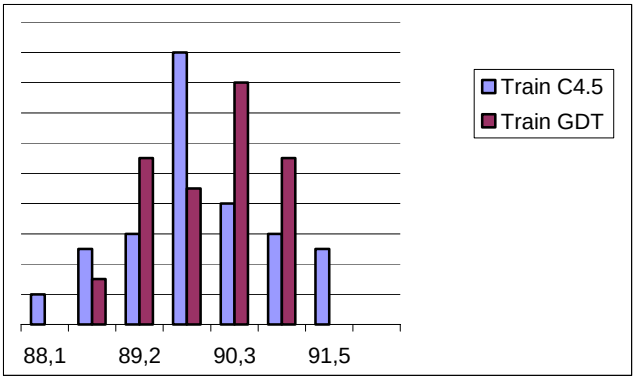
USER_ID	Sessioni	Durata_Visita	N°Pagine	Tempo_Pagina	Ora	Giorno	Sequenza	FEDELTA'
00001	38.74	509.66	18.16	1164.63	2	4	8	1
00002	11.07	359.57	42.03	580.99	2	4	8	0
00003	35.07	984.22	28.85	1126.62	4	3	1	1
00004	33.75	12.92	5.68	621.07	4	6	10	0
00005	35.74	937.47	39.39	945.67	4	5	2	1
00006	47.52	530.80	17.20	425.43	3	4	9	1
00007	15.81	197.51	48.29	1077.85	5	7	3	1
00008	20.14	547.64	43.41	61.08	4	6	5	0
00009	29.18	10.27	22.38	79.30	4	3	10	1
00010	25.40	114.31	3.69	1112.76	5	3	6	1
00011	13.44	624.29	8.69	540.67	5	3	4	1
00012	33.23	361.46	26.78	1144.58	4	2	10	0
00013	37.10	324.55	38.66	36.89	3	1	8	1
00014	12.76	287.06	47.36	927.16	2	3	2	0
00015	32.18	420.02	6.88	738.59	2	5	1	1
00016	25.45	697.07	6.90	155.86	2	5	2	1
00017	14.07	1181.57	4.61	133.20	3	6	9	0
00018	34.02	184.11	14.54	96.05	4	7	6	0
00019	41.11	957.78	27.58	938.40	3	1	3	0
00020	16.33	781.16	39.00	220.76	5	6	6	1
00021	26.44	231.23	43.87	701.98	5	4	10	1
00022	22.02	562.95	44.50	1177.22	4	3	5	0
00023	41.53	385.65	11.19	615.82	4	4	8	1
00024	21.17	266.75	1.37	776.07	1	2	3	0

Nell'impostazione dei parametri dell'algoritmo, dopo aver sperimentato differenti alternative, si è ritenuto proficuo attenersi a valori ritenuti oramai "classici" (Koza 1992):

Dimensione della popolazione:	500
Numero di generazioni:	51
Massima profondità dell'albero:	12
Tasso di selezione:	0.1
Tasso di crossover:	0.2
Tasso di mutazione:	0.01

L'algoritmo viene addestrato e successivamente verificato per cinquanta volte su differenti campioni di 2000 visitatori (50% di ciascun campione viene utilizzato per il training ed il restante 50 % per il testing); lo stesso tipo di esperimento è stato condotto con l'algoritmo C4.5 i cui principi di funzionamento sono stati precedentemente delineati (cfr. pagg. 2,3).

Il grafico che segue rappresenta la distribuzione del tasso di correttezza (corrispondente in GDT al valore di fitness (2)) su cinquanta prove di addestramento effettuate con GDT e C4.5:

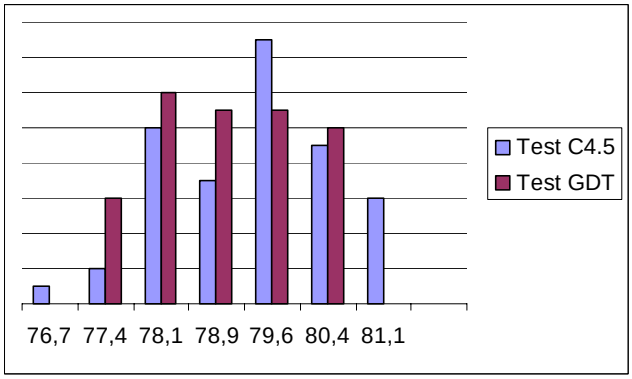


Train C4.5
Media = 92.24
Dev.St.= 0.98

Train Agents
Media = 90.01
Dev.St.= 0. 84

È riscontrabile in fase di addestramento una lieve predominanza nella prestazione media dell'algoritmo C4.5 rispetto a GDT compensata da una minore dispersione dei risultati di quest'ultimo intorno alla media.

La medesima situazione si ripropone per la porzione out-of-sample: ovviamente entrambe le tecniche fanno registrare un degrado nel proprio potere classificatorio pur mantenendosi su livelli di predittività molto interessanti:



Test C4.5
Media = 83.80
Dev.St.= 1.29

Test Agents
Media = 78.98
Dev.St.= 1. 11

È lecito attendersi che questa differenza di comportamento, peraltro confermata in altre sessioni sperimentali, sia da imputare alla diversa genesi e conformazione degli alberi prodotti dai due metodi; se ad esempio osserviamo un costrutto decisionale in cui si è sviluppata una delle soluzioni di GDT, possiamo riscontrare una morfologia "disordinata" che mai avremmo potuto ottenere con C4.5:

```
SE Pagg/Sessione < 6.89
  SE Durata sessione < 164 ALLORA Assidui = 129/138
ALTRIMENTI SE Durata sessione > 317 ALLORA Assidui =
522 / 539
ALTRIMENTI SE Pagg/ Sessione > 16 ALLORA Assidui = 3
/ 3
   ALTRIMENTI SE Fascia oraria = 4
  SE Sequenza = 4 ALLORA Assidui = 17 / 21
  ALTRIMENTI Sporadici = 271 / 299
```

A ben vedere infatti la classificazione risultante non si snoda attraverso una progressiva capillarizzazione che passi da insiemi più popolosi ad insiemi sempre più ridotti – come avviene nel C4.5 - ma attraverso un flusso irregolare che pone in alternativa caratteri discriminanti indipendentemente dal numero di individui che, nelle differenti classi, posseggono tali caratteri; l'indagine sulla relazione tra struttura degli alberi risultanti e potere descrittivo degli stessi costituisce interessante oggetto per ulteriori approfondimenti futuri.

V. Conclusioni

La conoscenza, e dunque la classificazione, del comportamento degli utenti internet è fattore irrinunciabile per l'adeguamento dell'offerta di servizi ad una domanda fortemente differenziata e volubile.

Tra gli strumenti di classificazione mirata multidimensionale che la soft computing mette a disposizione dell'esplorazione di ampie basi di dati i classificatori basati su agenti adattativi rappresentano una scelta interessante poiché forniscono in genere rappresentazioni sintetiche e leggibili, compatibili con le diffuse esigenze di integrazione tra modelli artificiali e la competenza dell'esperto umano.

Il punto di partenza del presente lavoro è stato pertanto l'intento di contribuire all'adozione degli agenti adattativi nell'analisi strategica dei siti web.

Punto di transito (non certo d'arrivo) è invece la proposta di ottimizzare le euristiche di partizione dei Decision Trees attraverso procedure evolutive al fine di evitare che suddivisioni locali, effettuate tenendo conto dell'efficacia su un singolo nodo, possano compromettere la correttezza della classificazione complessiva.

Il tratto che risulta emergere da una comparazione tra il sistema classificatore qui descritto e C4.5 è quello di un potere discriminante analogo sia in fase di addestramento che di testing con una tendenza più accentuata da parte degli agenti adattativi a contenere la dispersione dell'errore intorno alla media; il fondamento di tale tendenza nella particolare genesi e conformazione degli alberi in una procedura evolutiva è sicuramente uno degli aspetti da approfondire nel prosieguo della presente ricerca.

La possibilità per un sistema ad agenti di costituire per il futuro una valida alternativa a metodi deterministici dev'essere valutata anche alla luce dell'impegno computazionale che rispetto ad essi comporta; tuttavia nelle prove sin'ora affrontate l'indubbio onere di calcolo si è rilevato di entità relativa e completamente ammortizzato – rispetto ai tempi ed ai costi decisionali sostenibili in questo tipo di problematica – dalla potenza di calcolo attualmente disponibile.

VI. BIBLIOGRAFIA

[1] Armstrong R., Freitag D., Joachims T., and Mitchell T. "Webwatcher: A learning apprentice for the world wide web." In Proc. AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments. 1995

[2] Balabanovic M., Yoav Shoham, and Yun Y. "An adaptive agent for automated web browsing. Journal of Visual Communication and Image Representation." 6(4), 1995

[3] Brown C.M. , Danzig B.B., Hardy D., Manber U., and Schwartz M.F. "The harvest information discovery and access system. In Proc. 2nd International World Wide Web Conference, 1994".

[4] - Bot, M. "Application of Genetic Programming to the Induction of Linear Programming Trees" (Master thesis)

[5] - Bot, M.,"Improving Induction of Linear Classification Trees with Genetic Programming", Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000 Conference paper) p. 403-410

[6] Bala, J., W., Huang, J., Vafaie, H., Jong, K., A., D., and Wechsler, H., "Hybrid learning using genetic algorithms and decision trees for pattern-classification", proceedings of the fourteeth International joint Conference on Artificial Intelligence (Conference paper)

[7] Cantú-Paz, E. and Kamath, C., "Using Evolutionary Algorithms to Induce Oblique Decision Trees", Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000) p. 1053-1060

[8] Chang C. and Hsu C. "Customizable multi-engine search tool with clustering". In Proc. of 6th International World Wide Web Conference, 1997

[9] Doorenbos R.B., Etzioni O., and Weld D.S. "A scalable comparison shopping agent for the world wide web". Technical Report 96-01-03, University of Washington, Dept. of Computer Science and Engineering, 1996.

[10] Frakes W. B. and Baeza-Yates R. Information Retrieval Data Structures and Algorithms. Prentice Hall, Englewood Cliffs, NJ, 1992.

[11] Fu, Z., "A Computational Study of Using Genetic Algorithms to Develop Intelligent Decision Trees", proceedings of the 2001 Congress on Evolutionary Computation CEC2001, p. 1382-1387

[12] Goldberg, D.E.: Genetic Algorithms in Search Optimization and Machine Learning - Addison-Wesley 1989

[13] Hajek S. "Il commercio delle garanzie: conoscenza del cliente internet mediante Genetic Classification Trees", Atti del convegno Il rischio di credito e le implicazioni di Basilea 2 Siena, 8-9 Marzo 2002.

[14] Holland, J. H. Adaptation in Natural and Artificial Systems, U. of Michigan Press, 1975

[15] Holland, J.H. Hidden Order, Helix Books, 1995

[16] Janikow, C., Z "A Genetic Algorithm for Optimizing Fuzzy Decision Trees", Proceedings of the Sixth International Conference on Genetic Algorithms (Conference paper), p. 421-428

[17] Kennedy, H., C., Chinniah, C., Bradbeer, P. and Morss, L., "The construction and evaluation of decision trees: a comparison of evolutionary and concept learning methods", Proceedings of the Evolutionary Computing on AISB International Workshop p. 147-161

[18] Kennard, D., L., "Using genetic algorithm and decision trees to produce a hybrid classification system" Genetic Algorithms at Stanford, 1995

[19] Koza, J.R. Genetic Programming: On the Programming of Computers by Means of Natural Selection - MIT Press 1992

[20] Miller, M.R. Computer aided financial analysis – Addison Wesley 1990

[21] Oostendorp K. A., Punch W. F., and. Wiggins R. W. "A tool for individualizing the web." In Proc. 2nd International World Wide Web Conference, 1994

[22] Pazzani M., Muramatsu J, and Billsus D. "Syskill & webert: Identifying interesting web sites." In Proc. AAAI Spring Symposium on Machine Learning in Information Access, Portland, Oregon, 1996

[23] Pyle, D., Data Preparation for Data Mining – Morgan Kaufmamm Publishers - 1999

[24] Rouwhorst, S., E. and Engelbrecht, A., P. "Searching the Forest: Using Decision Trees as Building Blocks for Evolutionary Search in Classification Databases", Proc. of the 2000 Congress on Evolutionary Computation. p. 633-638

[25] Quinlan, J.R C4.5: programs for Machine Learning – Morgan Kaufmann Publishers 1993

[26] Ryan, M., D. and Rayward-Smith, V., J., "The Evolution of Decision Trees", Genetic Programming 1998: Proceedings of the Third Annual Conference p. 350-358

[27] Shirasaka, M., Zhao, Q., Hammami, O., Kuroda, K. and Saito, K., "Automatic Design of Binary Decision Trees Based on Genetic Programming", Second Asia-Pacific Conference on Simulated Evolution and Learning

[28] Spertus E. "Parasite: mining structural information on the web". In Proc. of 6th International World Wide Web Conference, 1997.

[29] Tanigawa, T. and Zhao, Q., "A Study on Efficient Generation of Decision Trees Using Genetic Programming" Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000 Conference paper), p. 1047-1052

[30] Vere, S., A., "Genetic classification trees", Evolutionary Algorithms in Management Application, 1995 (Conference paper) p. 277-289

[31] Yoshida, K., Yamamura, M. and Kobayashi, S., "Generating Pareto Optimal Decision Trees by GAs", proceedings of the 4th International Conference on Soft Computing (Conference paper), p. 854-859.