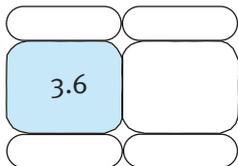




INTRODUZIONE ALLE RETI NEURALI ARTIFICIALI

Marco Gori

Nonostante gli straordinari successi dell'elaborazione dell'informazione, che stanno esercitando un impatto di portata storica nella vita quotidiana, competenze percettive quali localizzare un oggetto in una scena, riconoscere la voce in ordinarie condizioni reali, prendere decisioni basate sul "senso comune", risultano ancora compiti estremamente difficili per le macchine. Nel seguito, viene presentato un quadro generale dell'elaborazione neurale: successi, fallimenti e prospettive applicative.



1. IL MONDO SOTTO-SIMBOLICO

Gli odierni sistemi di elaborazione dell'informazione hanno compiuto prodigi che sono sotto gli occhi di tutti. Le macchine hanno automatizzato perfettamente processi considerati tipicamente di pertinenza umana, quali recuperare informazione in un archivio ed eseguire calcoli. Con l'intelligenza artificiale si sono spinte verso l'automazione del ragionamento simbolico, fino ai sistemi esperti, in grado di modellare e rendere fruibile la conoscenza di esperti in specifici settori. Ma nonostante i formidabili risultati conseguiti nell'automazione di alcuni processi intelligenti, soprattutto di alto livello, le macchine offrono ancora un comportamento piuttosto primitivo e incomparabile con l'uomo nella simulazione della maggioranza dei processi percettivi. La difficoltà di automatizzare tali processi è spesso trascurata per il fatto che già nel mondo animale sono presenti capacità percettive talvolta straordinarie. Tali capacità, sviluppate in secoli di processi evolutivi, risultano difficili da replicare usando i modelli di cal-

colo simbolico alla base degli attuali elaboratori.

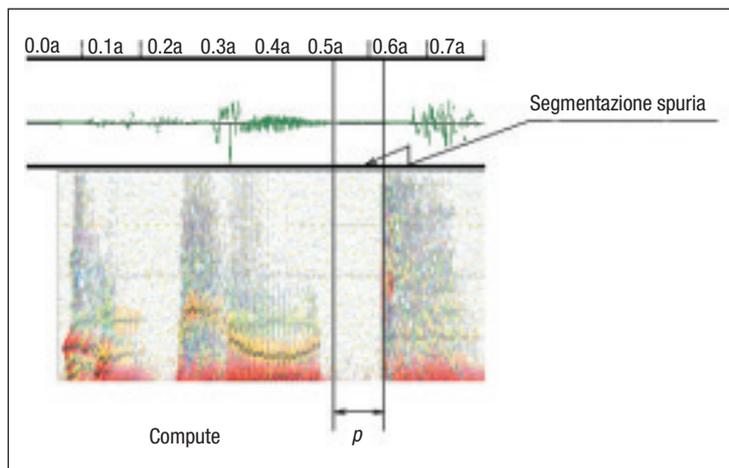
Si consideri, per esempio, il problema della descrizione dell'informazione presente in un'immagine che richiede la localizzazione e il riconoscimento di oggetti significativi per l'uomo. Tale processo richiede la capacità di segmentazione che, tuttavia, non può aver luogo solo con operazioni di basso livello basate, per esempio, sul rilievo di variazioni di luminosità. La segmentazione in una scena non può prescindere da processi cognitivi in grado di esibire competenza sugli oggetti e, più generalmente, del mondo oggetto dell'elaborazione. Inoltre, soprattutto nel mondo tridimensionale, gli oggetti si propongono con molteplici viste e, di nuovo, la loro percezione, sembra richiedere modelli ben più sofisticati di semplici comparazioni con oggetti di un dizionario. L'analisi attenta della nozione di similarità di oggetti o quotidiane acquisizioni quali, per esempio, "la facciata di una casa" (Figura 1) indicano lo sviluppo di competenze che non sono basate su elaborazione simbolica e ragionamenti qualitativi di dif-

ficile formalizzazione. L'esempio indicato in figura 1 illustra, inoltre, un livello di sofisticazione dell'analisi umana delle scene che risulta molto difficile da trasferire alle macchine: con uno sguardo attento si riconosce facilmente che le case nella scena sono, in realtà, miniature. La foto ritrae un paesaggio di un piccolo paese della montagna pistoiese in occasione del Natale, ma discernere il mondo artificiale miniaturizzato dal mondo reale è un compito realmente arduo per le macchine. Il problema del riconoscimento automatico della voce offre difficoltà simili. Il segnale vocale rappresentato in figura 2, assieme al suo spettrogramma, illustra alcuni aspetti della difficoltà del problema. Si presenta di nuovo il problema della segmentazione; anche in questo caso non si può fare affidamento su elementari elaborazioni di basso livello del segnale, quale per esempio il controllo del livello per separare le parole. Infatti, basta per esempio la presenza di occlusive sorde all'interno di una parola per il fallimento della segmentazione; la parola *compute* (si veda, a tal proposito, il segnale corrispondente a una sua pronuncia in figura 2) verrebbe, infatti, spezzata in due parti per la presenza dell'occlusiva sorda "p". Per via dell'enorme variabilità dovuta alla velocità di pronuncia, alla prosodia, al parlatore e a varie altre condizioni di rumore, le parole, inoltre, non sono facilmente rappresentabili mediante un dizionario di centroidi, ovvero di "istanze medie" di riferimento delle parole del dizionario.

I problemi menzionati si presentano ormai in molteplici applicazioni nelle quali l'elaborazione di informazione multimediale assume un ruolo sempre più rilevante. Si pensi, per esempio, alla navigazione autonoma di un *robot*, ai *data base* visuali, alla definizione di interfacce personalizzate, alla gestione di immagini di documenti, alla concezione di modelli per l'estrazione di informazione dal *web*. I problemi menzionati, assieme ad altri, hanno in comune il fatto che non sembrano naturalmente affrontabili mediante soluzioni basate su elaborazione simbolica. L'informazione da elaborare si presenta con una codifica a cui non è semplice attaccare significato. Il prodigio della scienza dei calcolatori proviene tipicamente dalla conoscenza del si-



FIGURA 1
Un paesaggio della montagna pistoiese durante il Natale



gnificato degli ingressi e dalla loro conseguente elaborazione mediante algoritmi. Non è, tuttavia, sempre possibile, o comunque verosimile, associare agli ingressi un significato e una conseguente caratterizzazione simbolica¹. Serve, dunque, dotare i calcolatori di processi computazionali che non siano necessariamente basati sulla metafora dell'algoritmo, secondo cui la soluzione di un problema avviene mediante un processo costruttivo atto ad esplicitare, simbolicamente, le elaborazioni sugli ingressi caratterizzati simbolicamente. Per alcuni problemi, questo approccio non appare naturale e risulta di difficile, se non impossibile, formalizzazione.

FIGURA 2
Segnale e spettrogramma relativo alla pronuncia del verbo "to compute"

¹ Nelle scienze cognitive questo problema è noto come *the symbol ground problem*.

2. LA METAFORA NEUROBIOLOGICA

Allo stato attuale, a differenza delle macchine, l'uomo è un ottimo esempio di "sistema" in grado di elaborare informazione sotto-simboliche. Tali elaborazioni, come ogni altro processo cognitivo, hanno sede nel cervello, una complessa struttura neurobiologica, attualmente decifrata in modo piuttosto accurato per quanto riguarda gli aspetti anatomici. È noto che c'è un "mattoncino elementare" che caratterizza tutte le strutture cerebrali, una cellula, denominata *neurone*, che è sede di processi elettrochimici responsabili per la generazione di campi elettromagnetici. Come è illustrato in figura 3, i neuroni sono composti

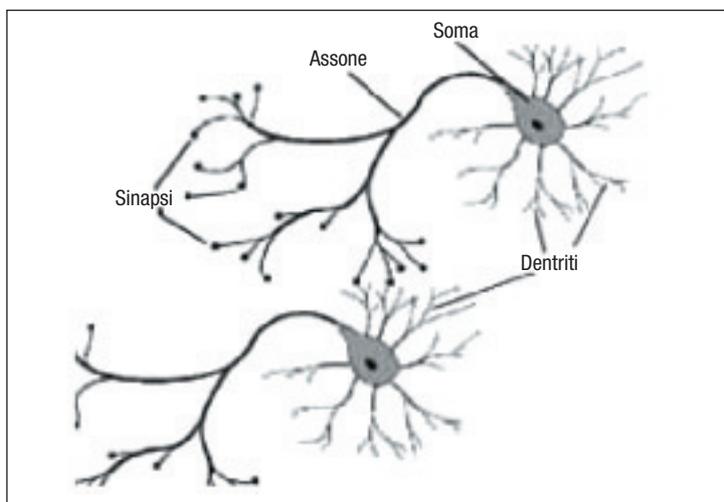


FIGURA 3
Neurone e sua struttura cellulare con soma, dentriti e connessioni sinaptiche

da un corpo detto *soma* e da due tipi di diramazioni: i dentriti e il cilindrase o assone. Nel cervello umano sono presenti tipicamente oltre 100 miliardi di neuroni, ciascuno interconnesso a circa altri 10.000. Nelle interconnessioni ha luogo la *sinapsi*, un processo elettrochimico atto a rinforzare o inibire l'interazione cellulare. I segnali rilevabili hanno un potenziale dell'ordine di alcune decine di millVolt e si presentano come treni di impulsi con frequenza intorno ai 100 Hz, con opportune modulazioni. Sono noti modelli sofisticati che esprimono il potenziale della cella (attivazione) in funzione del potenziale delle celle interconnesse. È opinione condivisa da ricercatori nel mondo delle scienze cognitive che i segnali elettrici presenti nei neuroni siano alla base dell'elaborazione dell'informazione a livello cerebrale. Le capacità cognitive sarebbero, dunque, in relazione all'elaborazione dei segnali presenti nei neuroni. Inoltre, c'è evidenza sperimentale per sostenere che la struttura cerebrale e le sinapsi siano influenzate dalla vita degli individui, dalle loro esperienze, dall'apprendimento di compiti specifici. È il particolare *pattern* di interconnessioni e la forza delle connessioni sinaptiche che definisce le proprietà funzionali di una particolare porzione del cervello. Si è, infatti, verificato sperimentalmente che le funzioni cognitive risiedono in particolari zone e che tali funzioni possono essere perse a seguito della "rottura" dei legami sinaptici ed

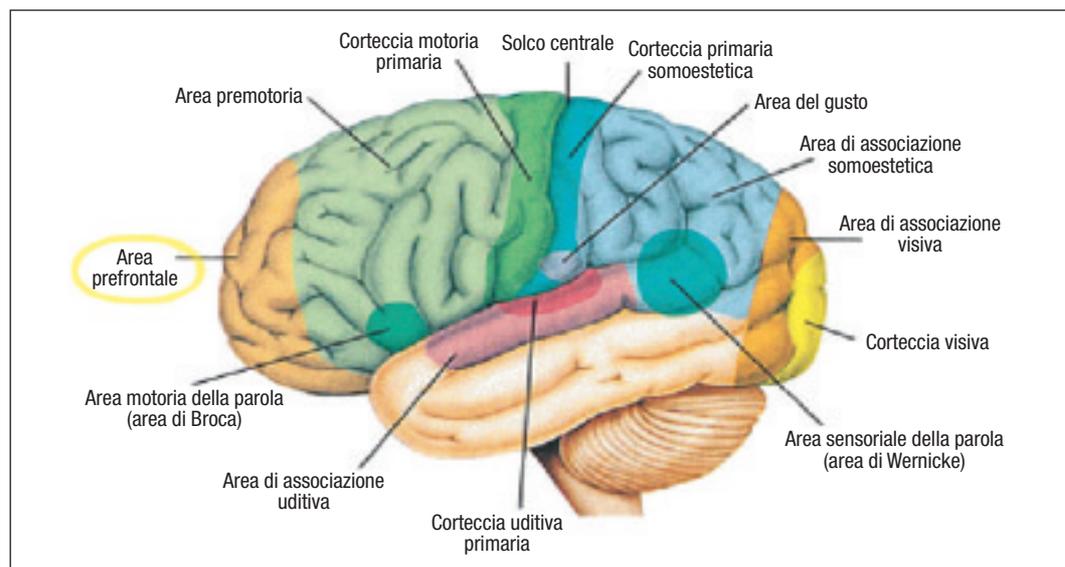


FIGURA 4
Organizzazione del cervello umano e sua localizzazione funzionale [13]



eventualmente recuperate, almeno in parte, con successivi processi di apprendimento atti a instaurare nuovi pattern di interconnessione sinaptica (Figura 4).

Dato che la struttura cerebrale e il comportamento elettromagnetico delle singole cellule neuronali sono noti, i ricercatori si sono ben presto chiesti se si possono operare induzioni sui comportamenti collettivi delle cellule neuronali, e dunque del cervello umano, e se si possono trarre utili suggerimenti e ispirazioni per la costruzione di macchine in grado di replicare compiti connotati da una forte componente di elaborazione sotto-simbolica, attualmente di difficile soluzione per i calcolatori. Il lavoro di McCulloch & Pitts [8] è forse il primo significativo passo in questa direzione, la prima analisi completa, soprattutto dal punto di vista formale, che fa intuire come semplici unità con sinapsi eccitatorie e inibitorie e con apposita soglia siano in grado, in virtù di un processo collettivo, di rappresentare complesse proposizioni. E questo sembra indurli a un certo ottimismo anche per le possibili implicazioni sulla comprensione dei processi cognitivi umani: "*Mind no longer goes more ghostly than a ghost*". Tuttavia il lavoro di McCulloch e Pitts sembra essere stato più rilevante per gli sviluppi nel settore dei calcolatori che non delle scienze cognitive. Carpire i segreti della mente dall'osservazione dell'attivazione cerebrale è una sfida affascinante, ma questo problema di *reverse engineering* sembra essere terribilmente intrappolato nella complessità del sistema neuronale umano. Si tratta sostanzialmente di indurre regolarità e leggi dall'osservazione, come in altri settori delle scienze. Questo problema possiede, tuttavia, un'infinità di sfaccettature e, soprattutto, richiede un processo di induzione che sembra inerentemente intrappolato nella complessità dei sistemi dinamici oggetto dello studio. Inferire regole dagli esempi sembra essere difficile anche in casi elementari; si pensi, a titolo di esempio, al problema dell'inferenza induttiva di grammatiche, che consiste nel determinare la grammatica che genera un linguaggio presentato mediante esempi. Sfortunatamente, questo problema è in-

trattabile perfino in caso di semplici grammatiche [3].

Era, tuttavia, ben chiaro ai padri dell'informatica che non è necessaria una perfetta emulazione dei processi neurobiologici per l'emergenza di capacità cognitive. Molti modelli connessionistici sono, infatti, solo ispirati dal paradigma biologico a livello di unità neuronale e si basano sulla struttura indicata in figura 3, dove si eredita il principio che l'attivazione neuronale (potenziale associato all'unità) è soggetta a eccitazioni e inibizioni dalle unità connesse. In particolare, l'attivazione dell'unità i dipende dall'attivazione della generica unità j mediante un parametro associato alla connessione tra le due unità, che modella il principio elettrochimico della sinapsi. In seguito, sarà illustrato come l'utilizzo di modelli di calcolo basati su reti neurali artificiali sia in grado di esibire quello che Lotfi Zadeh ha definito *softcomputing* secondo cui il requisito "trova sempre la soluzione esatta" diventa "trova spesso una soluzione approssimata".

La ricerca sulle reti neurali artificiali si è evoluta attraverso alterne vicende. Sin dagli albori dell'informatica, l'elaborazione basata su algoritmi e i modelli neurali centrati sull'apprendimento da esempi si sono sviluppati in parallelo. Verso la fine degli anni '60, Marvin Minsky e Simon Paper [10] pubblicano "Perceptrons", un libro che analizza con grande lucidità ed elegante formalizzazione le capacità computazionali del perceptrone di Rosenblatt. La comunità scientifica recepisce principalmente l'analisi critica del perceptrone e segue una fase di stagnazione che si protrae fino agli inizi degli anni '80. L'interesse rifiorisce, in particolare, per i lavori di Hopfield e del *Parallel Distributed Research Center* sulle reti neurali multistrato con l'algoritmo di apprendimento *Backpropagation*. Altre tappe importanti della ricerca nel settore sono riassunte nella tabella 1.

3. ARCHITETTURE NEURALI

Le neuroscienze hanno permesso di stabilire che la struttura cerebrale è caratterizzata dalla presenza di cellule neuronali con comportamenti vari e, soprattutto, da pattern di interconnessioni neuronali diversi a secon-

TABELLA 1
Alcuni eventi significativi che hanno marcato la storia delle reti neurali artificiali

I era	Eventi significativi
1943	McCulloch and Pitts, formalizzazione del neurone artificiale [8]
1949	D. Hebb e l'apprendimento per auto-organizzazione [6]
1956	"Dartmouth Summer Research Project on AI" con (Minsky, McCarty, Rochester, Shannon)
1960	Widrow: ADALINE [14]
1962	Il perceptron di Rosenblatt [11]
1969	"Perceptrons", Minsky & Papert (edizione espansa [10])
70s	Periodo "buio": degni di nota gli associatori di Anderson, i modelli per apprendimento senza supervisione di Kohonen, gli studi di Grossberg
II era	Eventi significativi
1982	Reti di Hopfield: memorie associative e soluzione di problemi [7]
1986	PDP e diffusione di Backpropagation [12]
1987	La prima conferenza significativa dell'IEEE a San Diego (II era)
1989	I chip neurali si affacciano sul mercato: <i>Analog VLSI and Neural Systems</i> [9]
1990	J. Pollack e le reti neurali che elaborano strutture dati
1994	Prima Conferenza Mondiale sull'Intelligenza Computazionale (Orlando)
1994	Nasce il progetto NeuroCOLT (<i>Computational Learning Theory</i>)
2001	L'IEEE approva la creazione della " <i>Neural Networks Society</i> "

do del compito cognitivo. Per i modelli artificiali è stata seguita una metafora simile: sono stati studiati diversi tipi di neuroni e diverse architetture associandovi le modalità di elaborazione concepite per implementare un determinato compito cognitivo. In figura 5 sono illustrati i due tipici neuroni artificiali che risultano, attualmente, i più interessanti dal punto di vista applicativo. Il primo, denominato neurone sigmoidale, è l'evoluzione del perceptrone di Rosenblatt [11], in cui il processo di decisione ha luogo mediante una funzione a gradino, invece, della funzione sigmoidale illustrata. Questi neuroni si eccitano per punti che sono situati sopra il piano di separazione, dove si annulla l'attivazione, e si inibiscono per punti situati al di sotto. Le unità del secondo tipo si eccitano per punti contigui al centro ($w_{ia}; w_{ib}; w_{ic}$) e si inibiscono quando ci si allontana, con una velocità commisurata al parametro radiale σ_i . È interessante notare che nei neuroni biologici l'attivazione ha la tipica forma di "un treno" di impulsi. La dipendenza dalle connessioni sinaptiche illustrata in figura 5 ha

effettivamente una genesi biologica, ma l'attivazione deve interpretarsi come la frequenza di treni di impulsi più che come un valore assoluto di potenziale. Sono stati anche studiati modelli formali, denominati *spiking neurons* che producono, a differenza dei due precedenti tipi di neuroni, treni di impulsi come i neuroni biologici. Gli studi sulle aggregazioni di tali neuroni e sui modelli di apprendimento costituiscono un interessante settore di ricerca che, tuttavia, non ha, ad oggi, prodotto risultati applicativi paragonabili a quelli conseguiti con i modelli semplificati in figura 5.

3.1. I perceptroni

I singoli neuroni descritti in precedenza possono essere utilizzati per il calcolo di semplici predicati, ma non possono certamente soddisfare molte significative esigenze reali. Per esempio, è immediato verificare che i neuroni sigmoidali non possono calcolare tutte le funzioni booleane di due variabili. In particolare, le funzioni $x_1 \otimes x_2$ e $\bar{x}_1 \otimes x_2$ non sono linearmente separabili, mentre, come è illustrato in figura 6 A, lo sono tutte le 14 altre

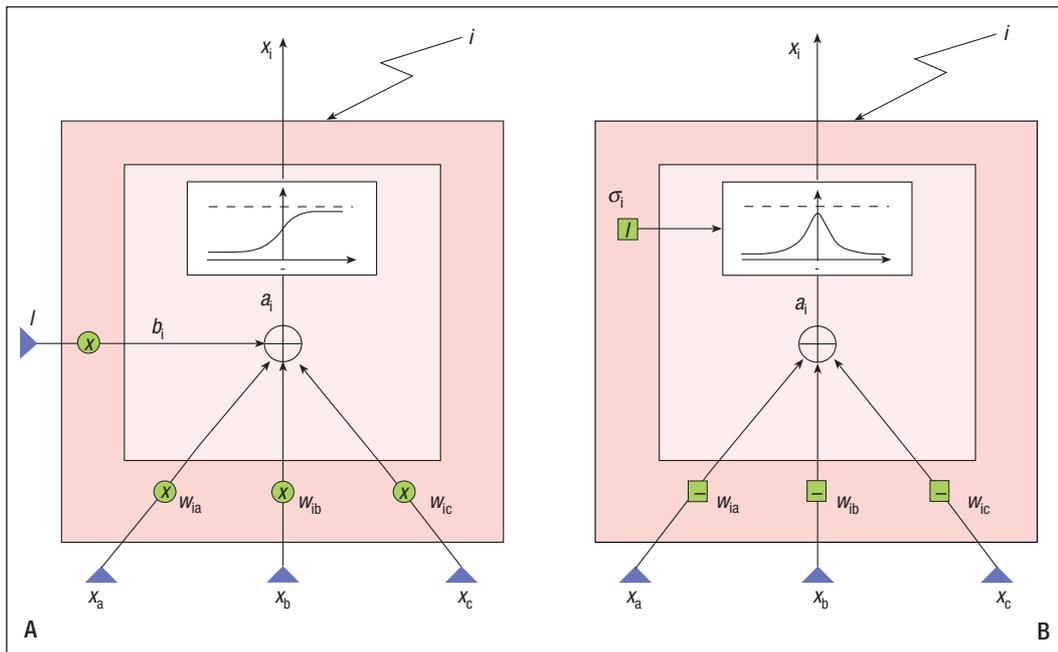
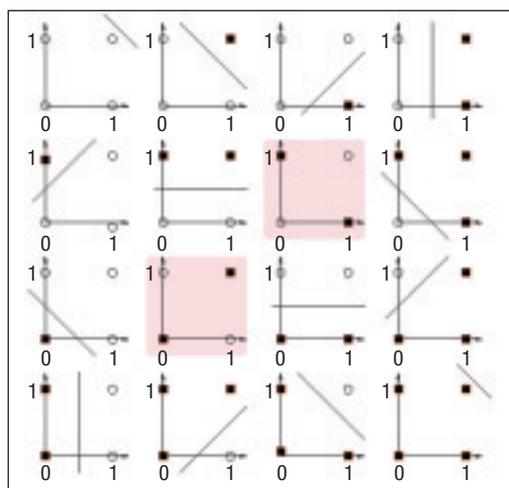


FIGURA 5
Due classici esempi di neuroni artificiali

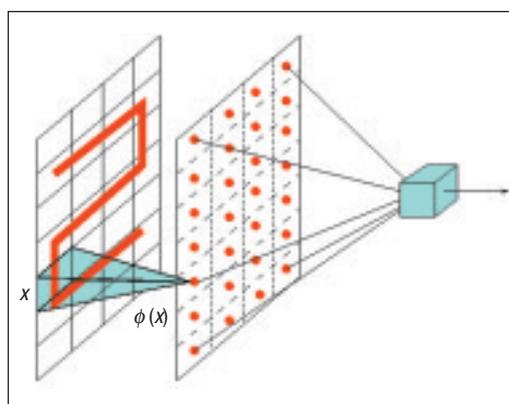
rimanenti. Più complesso è stabilire il comportamento del neurone di Rosenblatt nel caso di interessanti compiti cognitivi, quali il riconoscimento di forme.

Nella figura 6 B l'immagine è pre-elaborata mediante l'operatore di *pre-processing* $\Phi(x)$ che fornisce *feature* significative. Tale condizione impone, ovviamente, che il dominio dell'operatore sia limitato in modo da stabilire *feature* locali presenti nell'immagine indipendentemente da traslazioni e rotazioni. Indipendentemente dalla scelta dell'operatore, Minsky & Papert [10] hanno dimostrato che alcuni predicati topologici importanti, quali stabilire la connessione di una figura, non possono essere calcolati. Nella seconda metà degli anni '80, grazie soprattutto agli studi del *Parallel Distributed Processing research group*, sono state studiate, in modo sistematico, architetture neurali con architettura a grafo aciclico, in cui è definito un ordinamento parziale sui vertici. In tali architetture, un neurone può avere per genitori sia unità che ingressi (per esempio, il nodo 4, in Figura 7 A). Lo schema di calcolo si basa sulla "propagazione in avanti" delle attivazioni dei neuroni seguendo l'ordinamento parziale del grafo aciclico.

Per reti multi-strato lo schema di calcolo si riduce a una *pipe* sui livelli. Tali architetture erano, in realtà, già state concepite agli inizi



A



B

FIGURA 6
Due esempi delle limitazioni del perceptrone. A Funzioni booleane a due variabili; B Un perceptrone con il compito di stabilire se la figura elaborata è connessa

degli anni '60 e non differiscono sostanzialmente dallo schema illustrato in figura 6 B. Tuttavia, in quegli anni, si faceva riferimento a una sola unità neuronale in cui aveva luogo l'apprendimento e le altre unità erano semplicemente il risultato di una pre-elaborazione definita in modo esplicito, senza apprendimento. Nelle reti neurali di figura 7 le unità sono tutte uguali e sia la precedente elaborazione in avanti che l'apprendimento hanno luogo in modo omogeneo sui neuroni.

È stato dimostrato che le reti neurali *feed-forward* hanno un potere computazionale universale, ovvero sono in grado di calcolare ogni funzione di ragionevole interesse prati-

co con un numero sufficientemente grande di neuroni. Purtroppo, non sono disponibili concrete indicazioni di progetto, ma la proprietà di computazione universale può essere facilmente compresa almeno in due casi notevoli.

1. Funzioni booleane

È facile rendersi conto che le reti feedforward possono realizzare ogni funzione booleana. Infatti, ogni funzione booleana si può esprimere in prima forma canonica. A titolo di esempio, in figura 8 è illustrata la realizzazione della funzione XOR.

2. Funzioni di appartenenza

Una funzione di appartenenza è tale che $f_U(\mathbf{u}) = 1$ se e solo se $\mathbf{u} \in U$ [$f_U(\mathbf{u}) = 0 \Leftrightarrow \mathbf{u} \in \bar{U}$].

Per domini convessi (Figura 9) l'uscita si può determinare come AND di opportuni neuroni dello strato nascosto. Per domini non convessi (Figura 10) e/o concavi l'uscita si può determinare come l'OR di opportune unità nascoste (3 strati - eccetto l'ingresso). Si noti che questo metodo si basa su un processo costruttivo e che, dunque, non si può concludere che *servono necessariamente* due strati

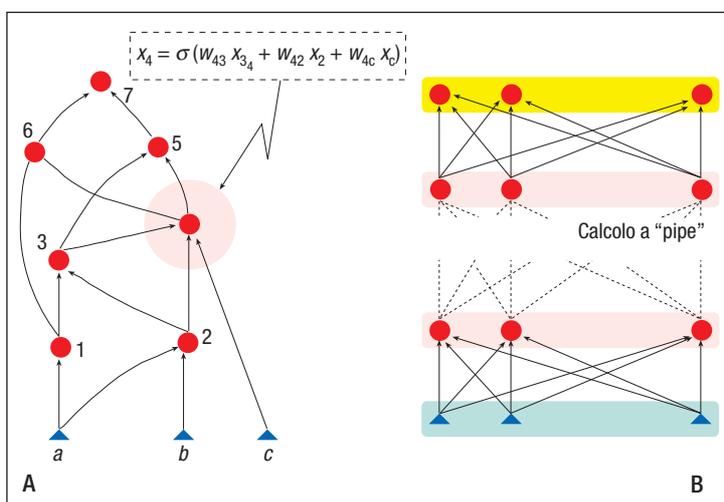


FIGURA 7

A Rete feedforward con struttura a grafo aciclico. B Rete multistrato

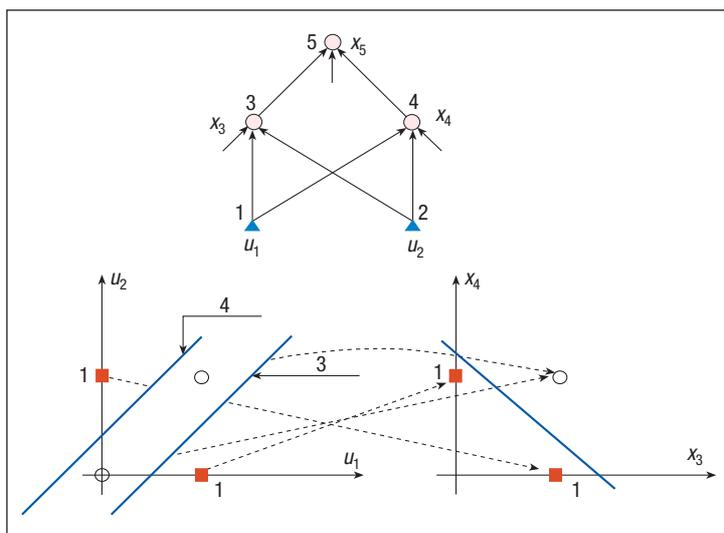


FIGURA 8

Realizzazione mediante perceptrone multistrato della funzione booleana XOR

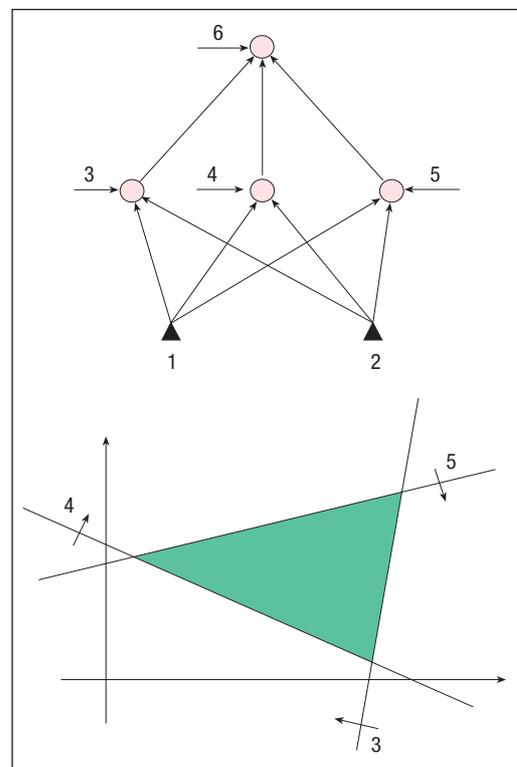


FIGURA 9

Tre unità nascoste che originano i tre iperpiani necessari per definire il dominio convesso

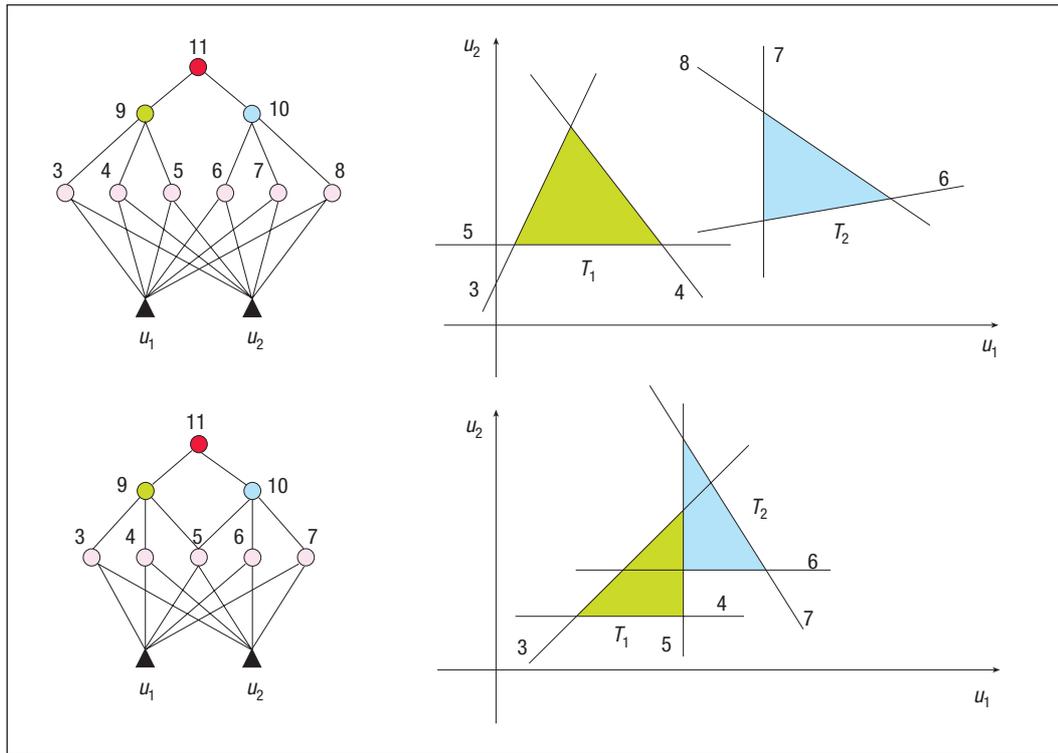


FIGURA 10
Domini non connessi possono essere ottenuti, per esempio, con due strati nascosti

nascosti per il calcolo di tali funzioni di appartenenza. Simili elaborazioni si possono eseguire anche usando neuroni a simmetria radiale di figura 5 B.

3.2 Reti neurali ricorsive

Le architetture neurali descritte in precedenza presuppongono schemi di “calcolo in avanti”, basati su un ordinamento delle unità. Come per le funzioni booleane, la presenza di cicli conduce a elaborazioni più complesse che coinvolgono sequenze e non singoli pattern. Una tipica situazione in cui risulta naturale un’elaborazione sequenziale è quella del riconoscimento di fonemi illustrata in figura 11.

L’elaborazione è sincronizzata in corrispondenza a ogni *frame*. I neuroni sono ancora del tipo illustrato in figura 5, ma oltre agli ingressi provenienti dal frame corrente, ai neuroni dello strato nascosto afferiscono anche come ingresso il valore delle uscite relative al frame precedente.

Le reti neurali ricorsive presentano strette connessioni con gli automi, ma il calcolo che ha luogo nelle unità neuronali ha natura continua e non discreta. In virtù di tale natura, le reti neurali ricorsive presentano anche una

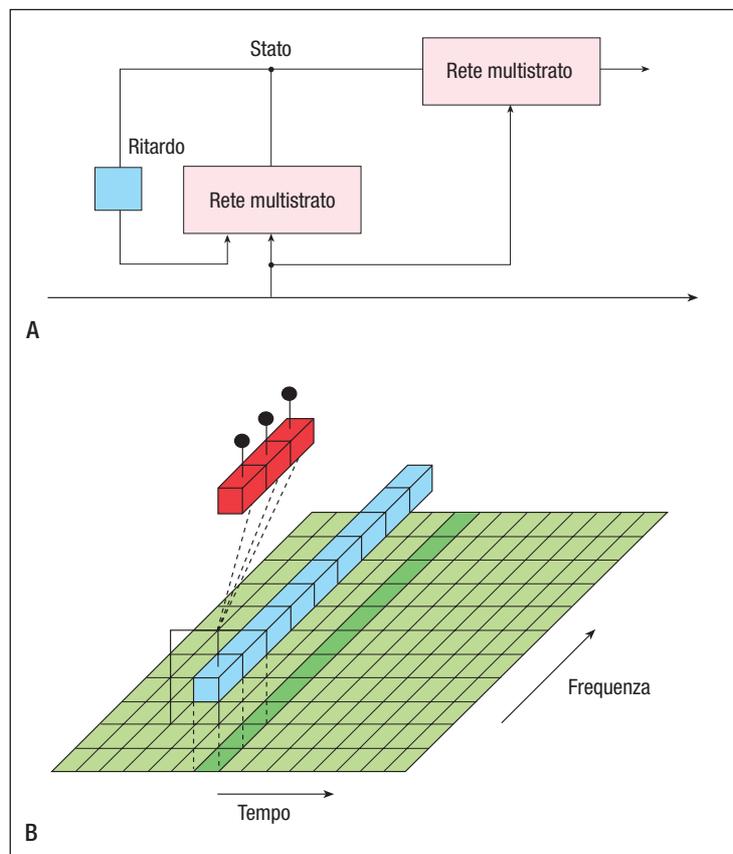


FIGURA 11
A Architettura di una generica rete ricorsiva. B Una rete neurale ricorsiva

forte connessione con i sistemi dinamici lineari, anche se la loro struttura dinamica è orientata a processi decisionali.

Le reti ricorsive non elaborano solo sequenze ma, direttamente, anche grafi i cui nodi contengono un vettore di numeri reali. Si potrebbe sempre ricondurre l'elaborazione di strutture a quella di opportune sequenze. Tuttavia, rappresentando un grafo come una sequenza si nascondono caratteristiche discriminanti per l'apprendimento. Inoltre, si può mostrare che la riduzione a lunghe sequenze derivante dalla codifica di strutture a grafo in stringhe rende il problema dell'apprendimento da esempi più costoso. L'elaborazione su grafi ha luogo estendendo il concetto di calcolo dello stato in un automa a stati finiti dal caso di sequenze a quello di alberi e, più generalmente, di grafi aciclici ordinati [5]. Per le architetture illustrate l'aggiornamento dello stato avviene in modo sincrono rispetto all'alimentazione di un nuovo ingresso della sequenza o della struttura dati. La struttura di una rete neurale ricorsiva può, tuttavia, operare anche elaborazioni sequenziali, mediante aggiornamento dello stato, di ingressi tenuti fissi. L'esempio più classico è quello delle reti di Hopfield, illustrate in figura 12. Si noti che l'uscita di ogni neurone è connessa a tutti gli altri e che non c'è connessione locale. Nell'esempio illu-

strato in figura un'immagine, che raffigura un pattern corrotto da rumore, viene presentata in ingresso alla rete ricorsiva. L'ingresso è costituito dai *pixel* dell'immagine o, più in generale, da una forma de-campionata a risoluzione più bassa dell'originale. Con opportune scelte dei pesi delle connessioni², mantenendo l'ingresso costante, la rete ricorsiva procede ad aggiornamenti delle attivazioni dei neuroni finché, dopo una fase di rilassamento, raggiunge un punto di equilibrio. Com'è illustrato in figura, il punto di equilibrio corrisponde all'immagine filtrata dal rumore. In pratica, una rete di Hopfield con N ingressi, permette in modo affidabile di memorizzare un numero di pattern intorno a $0,15N$ e può, pertanto, essere utilizzata come memoria associativa.

4. APPRENDIMENTO DA ESEMPI

Nelle reti neurali artificiali, le architetture illustrate nel paragrafo precedente, assieme ai corrispondenti schemi computazionali, sono di scarso interesse senza il paradigma centrale dell'apprendimento, che viene ispirato a corrispondente paradigma neurobiologico. Apprendere in una rete neurale artificiale corrisponde a modificare il valore dei pesi delle connessioni sinaptiche. Tale processo è influenzato dagli esempi che concorrono a sviluppare concetti. I dati e l'interazione con l'ambiente concorrono con diversi protocolli allo sviluppo di competenze cognitive. In particolare, si individuano tre diverse modalità di apprendimento a seconda del ruolo esercitato dal supervisore del concetto: l'apprendimento con supervisione, l'apprendimento con rinforzo e l'apprendimento senza supervisione.

4.1. Protocolli di apprendimento

Nell'apprendimento con supervisione e con rinforzo, la rete neurale deve sviluppare un concetto sulla base delle interazioni con un supervisore, che provvede a istruire la rete, fornendo informazioni sul concetto.

Si consideri, per esempio, il problema della

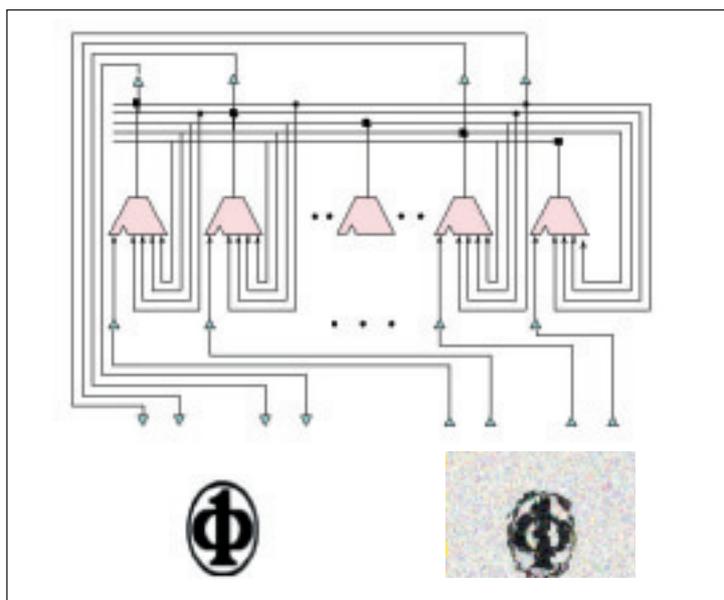


FIGURA 12
Rete di Hopfield utilizzata come memoria associativa per filtrare il rumore

² In particolare, la matrice delle connessioni è simmetrica.



classificazione di insetti illustrato in figura 13. La rete neurale esibisce la sua classificazione e interagisce con il supervisore che può fornire un'informazione completa o parziale sul concetto. Quando l'informazione è parziale si parla di "apprendimento con rinforzo"; tale informazione deve essere utilizzata nell'apprendimento per rinforzare comportamenti corretti e penalizzare quelli che originano errori.

Nell'"apprendimento con supervisione", invece, il supervisore fornisce l'informazione completa sul concetto, definendo, in questo caso, esattamente la classe di appartenenza. L'apprendimento di un concetto non richiede necessariamente l'interazione con un supervisore e può aver luogo anche mediante un'auto-organizzazione degli esempi. Apprendere senza supervisione significa aggregare esempi simili in regioni neuronali topologicamente vicine. In figura 14 è illustrata l'auto-organizzazione di esempi di classi diverse e l'eccitazione dei neuroni spazialmente correlati al concetto. Mentre per i due precedenti protocolli di apprendimento la variazione delle connessioni sinaptiche avviene cercando di ottimizzare l'errore rispetto all'informazione fornita dal supervisore: in questo caso, l'apprendimento è guidato da criteri di "similarità" nei dati.

In generale, i tre protocolli di apprendimento descritti sono formulabili come ottimizzazione di una funzione dei pesi della rete neurale. Nel caso dell'apprendimento con rinforzo e dell'apprendimento con supervisione, per rendere il comportamento della rete neurale conforme alla supervisione occorre minimizzare una funzione di errore che dipende dalla scelta dei pesi e misura l'errore rispetto alle informazioni del supervisore. Nel caso dell'apprendimento senza supervisione, l'auto-organizzazione per similarità dei dati può ancora, generalmente, formularsi come l'ottimizzazione di una funzione di armonia. Il problema di ottimizzare funzioni in grossi spazi è generalmente difficile per la potenziale presenza di minimi locali, che può rendere inefficaci le classiche euristiche di ottimizzazione basate sulla tecnica di massima discesa del gradiente.

Il corretto funzionamento di una neurale sull'insieme di apprendimento non offre, ovvia-

mente, garanzia di un altrettanto soddisfacente funzionamento su altri dati relativi allo stesso concetto, ma non utilizzati nella fase di apprendimento (insieme di test). Inoltre, è evidente che l'architettura della rete neurale gioca un ruolo fondamentale per l'efficienza della fase di apprendimento. Si consideri, ad esempio, il caso delle reti feedforward e il loro comportamento al variare del numero delle unità nascoste. In virtù della loro capacità universale di approssimazione, tali reti possono calcolare ogni concetto. Quando il numero delle unità nascoste cresce, non solo aumenta il potere computazionale, ma si può dimostrare che il problema della presenza dei minimi locali diventa progressivamente meno rilevante. Tuttavia, al crescere

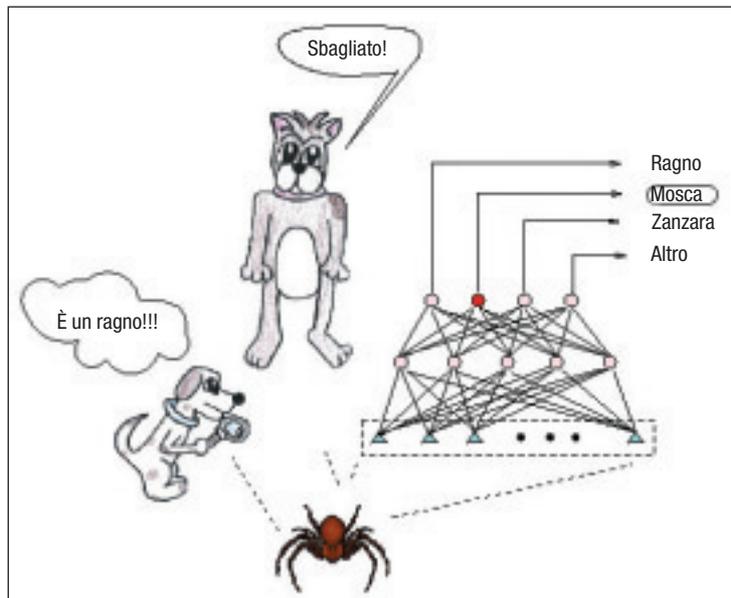


FIGURA 13
Il paradigma di apprendimento con supervisione e con rinforzo

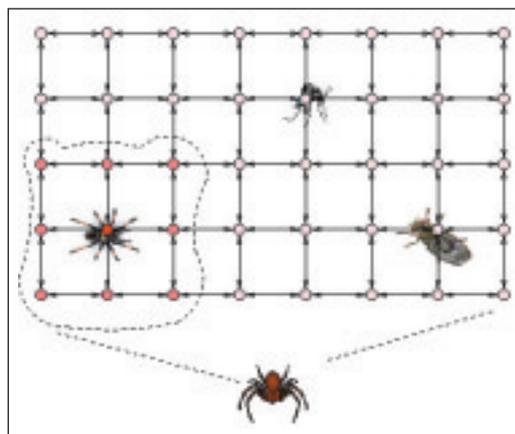


FIGURA 14
Apprendimento senza supervisione

della dimensione della rete la capacità di generalizzare su nuovi esempi tende a diminuire dato che il *fitting* sull'insieme di apprendimento ha luogo in un enorme spazio di parametri vincolati solo da pochi esempi. Questo origina una sorta di principio di indeterminazione dell'apprendimento secondo il quale non è possibile al variare dei pesi della rete neurale ottenere funzioni di errore senza minimi locali ed eccellente generalizzazione per nuovi esempi.

4.2. Algoritmi di apprendimento

La formulazione dell'apprendimento come ottimizzazione della funzione errore rispetto alla supervisione permette di attingere all'enorme letteratura di analisi numerica per la ricerca degli algoritmi più opportuni. Tuttavia, dato che in pratica si opera con reti neurali che possono raggiungere centinaia di migliaia di variabili³, si restringe di solito l'attenzione all'uso dell'euristica di massima discesa del gradiente che, essendo una tecnica del primo ordine permette di limitare spazio e tempo di calcolo. Algoritmi di apprendimento direttamente basati su tecniche iterative, come il gradiente, prendono il nome di algoritmi di tipo *batch*. In tali algoritmi, la variazione dei pesi avviene solo dopo aver elaborato tutti gli esempi dell'insieme di apprendimento. Si possono, però, concepire algoritmi nei quali i pesi sono aggiornati in corrispondenza della presentazione di ogni esempio (algoritmi di tipo *on-line*).

L'entità di variazione dei pesi in corrispondenza degli esempi può condurre, in questo caso, a enfatizzare il comportamento corretto sugli "ultimi" esempi presentati, dimenticando i vecchi esempi su cui la rete aveva appreso. È evidente che tale entità deve anche essere commisurata alla numerosità dell'insieme di apprendimento. Un aspetto particolarmente rilevante degli algoritmi di apprendimento, che è talvolta sottovalutato, è costituito dalle proprietà di località spaziale e temporale che sono tipicamente gradite per

motivi di efficienza computazionale. Si richiede, in sostanza, che lo schema di aggiornamento dei pesi preveda, per ogni neurone, l'uso di informazione disponibile per mezzo delle sole unità che sono direttamente collegate (località spaziale) e che tale informazione sia riferita solo all'istante di tempo precedente (località temporale).

Per reti ricorsive con architettura generica risulta difficile concepire schemi di apprendimento in grado di coniugare entrambe queste proprietà, mentre questo è possibile per architetture speciali, quali ad esempio quella di figura 11 B, in cui le connessioni che producono la ricorsività sono solo locali ai neuroni. Nei protocolli di apprendimento, considerati fino a questo punto, si è assunto che l'apprendimento consiste solo nella variazione dei pesi e che questo ha luogo a partire da una rete neurale con architettura predefinita. Il principio di indeterminazione, precedentemente menzionato, suggerisce, tuttavia, che la definizione stessa dell'architettura possa ragionevolmente costituire oggetto del processo di apprendimento. Tale assunzione, che ha solide basi neurobiologiche, conduce dunque allo studio di algoritmi atti a creare e cancellare connessioni sinaptiche oltre che a variarne il peso corrispondente. Sono stati concepiti algoritmi di *growing* e di *pruning* delle connessioni basati tipicamente sul principio della sensibilità dei pesi rispetto al comportamento della rete. Sono stati anche proposti algoritmi genetici per sviluppare l'architettura adeguata a un certo "task". Infine, la supervisione può essere fornita in modo più sofisticato mediante uno schema di insegnamento, che mira a presentare il desiderato *target* in modo progressivo.

5. PROBLEM SOLVING

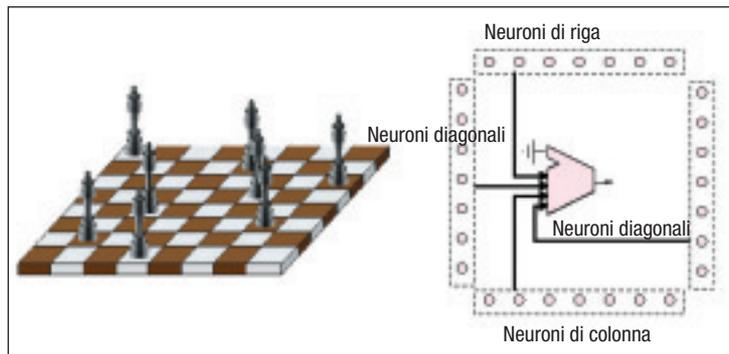
L'architettura ricorsiva della rete di Hopfield descritta nel paragrafo 3 è stata oggetto di molti studi non soltanto per le sue proprietà di memoria associativa, ma anche perché ben si presta alla soluzione di molti interessanti problemi di soddisfacimento di vincoli, che hanno spesso natura combinatoriale. Per illustrare questa interessante proprietà si consideri il classico problema di scacchi di al-

³ In alcuni esperimenti di riconoscimento vocale, nel gruppo di ricerca di Herve Bourlard (IDIAP, Svizzera) sono state utilizzate reti neurali con circa un milione di pesi.



locare 8 regine su una scacchiera in configurazione di non attacco, ovvero in modo tale che non si mangino⁴ (Figura 15).

Questo è un problema di soddisfacimento di vincoli. Occorre, infatti, che, per ogni regina, non siano presenti regine sulla stessa riga, la stessa colonna e le due diagonali. La soluzione del problema si ottiene mediante una rete ricorsiva con un numero di neuroni pari al numero di caselle della scacchiera. L'eccitazione di un neurone corrisponde alla presenza della regina sulla casella, l'inibizione corrisponde, invece, alla casella vuota. Per risolvere il problema occorre tradurre i vincoli del problema in corrispondenti vincoli sul valore delle attivazioni dei 64 neuroni. Le connessioni della rete neurale, solo inibitorie, si costruiscono associando a ogni neurone un peso negativo proveniente dai neuroni associati alle caselle che si trovano sulla stessa riga, sulla stessa colonna e sulle stesse due diagonali cui appartiene la casella associata al neurone in oggetto. Non ci sono, dunque, auto-connessioni e, inoltre, la matrice dei pesi è simmetrica; se l'unità i è connessa all'unità j da un peso w_{ij} allora vale anche il viceversa, cioè $w_{ji} = w_{ij}$. Si noti che a differenza del caso in cui la rete di Hopfield opera da memoria associativa, in questo caso non ci sono ingressi collegati e che la codifica del problema è tradotta nel pattern di interconnessioni. Si può dimostrare che con un simile insieme di collegamenti, partendo da una qualunque configurazione iniziale, la dinamica della rete neurale evolve verso un punto stabile in cui la soluzione rappresenta configurazioni con regine in posizione di "non attacco". Tuttavia, partendo da una configurazione casuale, l'evoluzione della dinamica della rete non garantisce che tutte le 8 regine siano piazzate sulla scacchiera. Si può anche dimostrare che l'evoluzione della dinamica corrisponde alla minimizzazione di una funzione energia e che i suoi minimi globali corri-



spondono a soluzioni del problema. Come nel caso dell'apprendimento tuttavia, la funzione può essere popolata da minimi locali, offrendo pertanto soluzioni spurie. La soluzione sommariamente illustrata per il problema delle 8 regine può essere estesa con metodologie generali per risolvere generici problemi di soddisfacimento di vincoli, tipicamente molto complessi dal punto di vista computazionale quali, per esempio, il problema del commesso viaggiatore e il *knapsack*. Le soluzioni offerte da questo approccio sono estremamente efficienti oltre a permettere una computazione parallela per ogni passo del processo dinamico. Il problema fondamentale, tuttavia, è che, come per l'apprendimento, si hanno talvolta soluzioni sub-ottime che possono non risultare soddisfacenti. In sostanza, con soluzioni basate su reti di Hopfield, una volta "programmate" le connessioni per codificare il problema da risolvere, si può anche conseguire in modo efficiente una soluzione per problemi intrattabili, ma questo non è ovviamente garantito⁵.

FIGURA 15

Il problema delle 8 regine e la sua soluzione mediante una rete di Hopfield

6. LE APPLICAZIONI

Uno dei motivi del successo delle reti neurali artificiali è probabilmente da ricercarsi nel loro massiccio utilizzo in innumerevoli applicazioni. Il paradigma di apprendimento da esempi su cui si basano permette, infatti, di affrontare problemi di natura anche molto di-

⁴ La generalizzazione di questo problema al caso di N regine è stato per anni oggetto di congetture. Si riteneva si trattasse di un problema computazionalmente intrattabile, ma a metà degli anni novanta si è dimostrato che esiste una soluzione polinomiale per la determinazione di una configurazione.

⁵ Si tratta, in sostanza, di una delle caratteristiche fondamentali del softcomputing menzionata nel paragrafo 2.

versa e di fornire soluzioni con uno sforzo relativamente limitato.

Questo è anche stato reso possibile dalla grande diffusione di pacchetti *software* per la simulazione dei modelli più importanti. I principali modelli neurali sono oggi disponibili anche in molti *tool* per *data mining* disponibili nei principali sistemi per basi di dati quali il DB2 (Database2). Oltre alla simulazione software, sono state studiate diverse soluzioni per l'implementazione in *hardware* di architetture neurali e dei corrispondenti algoritmi di apprendimento. Molti studi si sono concentrati su come utilizzare gli attuali modelli di calcolo parallelo per l'implementazione dello schema neurale, intrinsecamente parallelo. Sono fiorite innumerevoli soluzioni nei laboratori di ricerca che hanno avuto anche un certo impatto commerciale permettendo lo sviluppo di acceleratori neurali per integrare le capacità di calcolo di elaboratori tradizionali. Tali acceleratori sono tipicamente gestite da alcuni simulatori commerciali. L'impressionante evoluzione dei microprocessori che ha avuto luogo anche negli anni '90 ha, tuttavia, sostanzialmente ridimensionato l'importanza di tali soluzioni.

Si è anche assistito alla nascita di *chip* neurali analogici in grado di implementare i paradigmi di calcolo direttamente con variabili analogiche, senza bisogno di codifica discreta. In particolare, è degno di nota l'INTEL 80170, sviluppato nei laboratori INTEL all'inizio degli anni novanta. Studi simili sono stati compiuti soprattutto da Synaptics, (Object Recognizer Chip) e, in Italia, (TOTEM) della NeuriCam.

Uno dei problemi che ha, tuttavia, limitato lo sviluppo di chip tipo l'INTEL 80170 è la limitata precisione disponibile, che costituisce un problema soprattutto per gli algoritmi di apprendimento.

6.1. Applicazioni al riconoscimento di forme

Per illustrare la metodologia alla base di molte delle applicazioni riportate in tabella 2, si consideri il caso del riconoscimento di simboli grafici, eventualmente corrotti da rumore. Occorre pre-elaborare il pattern in modo da fornirne una rappresentazione più

compatta da utilizzare in ingresso alla rete neurale. La limitazione del numero degli ingressi risulta particolarmente importante per limitare il numero degli esempi necessari per una corretta generalizzazione delle reti a nuovi esempi.

Nella figura 16, è illustrato l'uso di un perceptrone multistrato per la classificazione di logo aziendali in 4 categorie. La rete ha 256 ingressi e 4 uscite, codificate in modo esclusivo, ovvero (1; 0; 0; 0); (0; 1; 0; 0); (0; 0; 1; 0); (0; 0; 0; 1). Il numero di neuroni nascosti si determina per tentativi utilizzando un test di validazione statistica.

Il perceptrone multistrato dimostra eccellenti capacità di discriminazione di classi, ma non risulta efficace per attribuire un livello di confidenza nella sua decisione.

In altri termini, mentre molte applicazioni ne hanno dimostrato la grande efficacia nella discriminazione di classi note a priori, si è ormai accumulata evidenza sperimentale e supporto teorico per concludere che il perceptrone non è in grado di attribuire in modo affidabile un peso alle sue decisioni. Questo rende tale rete neurale usata come classificatore inadatta a problemi in cui è necessario un comportamento di reiezione di pattern che non appartengono alle classi pre-stabilite.

Sempre usando il perceptrone multistrato, si può ovviare a questo inconveniente mediante la configurazione ad autoassociatore illustrata in figura 17.

6.2. Sistemi ibridi

Molte delle applicazioni delle reti neurali a problemi reali richiedono un'opportuna organizzazione di sistema e non semplicemente l'utilizzo diretto dei modelli descritti in questo articolo. Per esempio, l'estrazione dell'informazione da una fattura acquisita mediante uno scanner richiede un opportuno sistema per la gestione documentale, dove le reti neurali possono giocare un ruolo strategico in alcune parti critiche.

A titolo di esempio, si consideri il problema del riconoscimento di targhe automobilistiche acquisite mediante ordinarie telecamere in ambiente autostradale. Tale applicazione è, per esempio, interessante per le società di gestione del traffico autostradale in corri-

Settore applicativo	Prodotto
Marketing	Airline Marketing Assistant, BehavHeuristics Inc Add-ins per Microsoft Excel, NeuroXL, 1998 AREAS, valutazione automatica immobili, HNC Software
Previsioni finanziarie	Neurodimension www.nd.com, 1991 NetProfit (profitmaker.con), Neur. Appl. Corp. Appl.
Optical Character Recognition	Audre Neural Network, Audre Rec. Systems Appl. OmniPage 6.0 and 7.0 Pro for Windows, Caere OmniPage 6.0 Pro for MacOS AnyFax OCR engine FaxMaster, Delrina Technology Inc. VeriFone Oynx, lettore di assegni, Synaptics
Riconoscimento caratteri manoscritti	QuickStroke, ric. caratt. cinesi, Synaptics Teleform: ric. caratteri per fax, Cardiff Software Application, 1991
Riconoscimento manoscritti on-line	Apple Newton 120, Apple Lexicus Longhand, Lexicus (Motorola)
Nasi elettronici	AromaScan electronic nose, AromaScan Bloodhound Electronic Nose, Bloodhound Sensors Ltd e-NOSE 4000 electronic nose, Neotronics Scientific
Controllo di qualità cibi	test qualità birra Anheuser-Busch
Bond portfolio management	Global Bond, Econostat Ltd.
Controllo frodi (assegni)	Dunn and Bradstreet
Controllo frodi (carte credito)	Falcon, HNC Software Nestor In.
Verifica firma	Check Signature Verification System, NeuroMetric Vision System Inc.
Gestione rischio	Colleague, Aquarius, HNC Software
Predizione del consumo elettrico	Bayernwerk AG Application
Controllo chip microelettr.	INTEL
Controllo qualità gomme	Dunlop
Cancellazione di eco	AT&T/Lucent
Riconoscimento di banconote	BANK, D.F. Elettronica
Riconoscimento di targhe	PLARE, Società Autostrade e DII (Università di Siena)

TABELLA 2

Una lista di applicazioni di reti neurali che hanno dato origine a prodotti apparsi sul mercato



FIGURA 16
Classificazione di loghi aziendali con un perceptrone multistrato

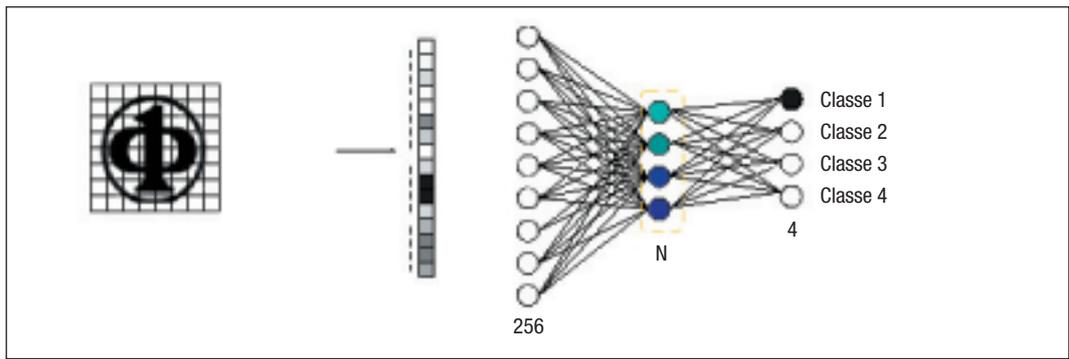
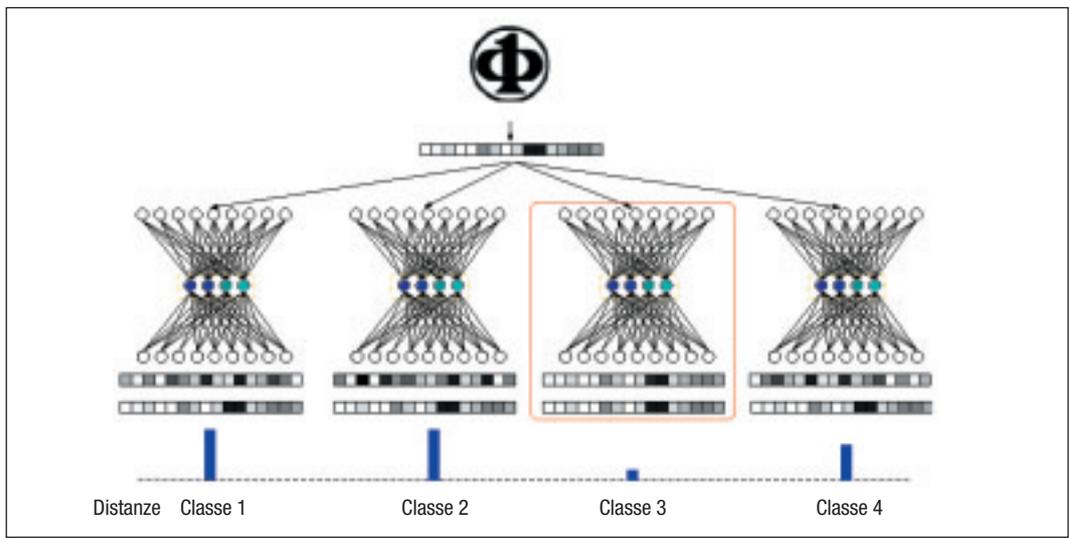


FIGURA 17
Autoassociatori neurali



spondenza delle stazioni di esazione a seguiti di infrazioni in impianti automatici. In figura 18, è illustrata l'architettura complessiva di un sistema per il riconoscimento di targhe in sperimentazione presso il Dipartimento di Ingegneria dell'Informazione dell'Università di Siena. Il sistema è composto da moduli sviluppati con tecnologia neurale (in rosa), da moduli basati su classici approcci di elaborazione delle immagini e da motori inferenziali. Un modulo di controllo provvede a sincronizzare le operazioni dei moduli *slave* delegati ad assolvere le funzioni di segmentazione della targa, dei caratteri e riconoscimento dei caratteri. Altri moduli esprimono vincoli grammaticali sulle stringhe possibili oltre a una probabilità a priori che si presenti una data targa. Il riconoscimento dei caratteri, che costituisce ovviamente l'attività critica, è basato su due moduli. Il primo contiene perceptron multistrato con struttura ad autoassociatore, che modellano le classi atte-

se. La struttura può integrarsi dinamicamente quando si presenta una eventuale altra classe. Tale modulo ha la funzione di stabilire una lista di classi candidate, mentre il modulo a fianco, basato su perceptron multistrato con struttura a classificatore, serve a raffinare la decisione. Tali classificatori sono tipicamente invocati dal modulo centrale quando i candidati si riferiscono a classi tipicamente molto confuse. In tal caso, appositi classificatori assolvono unicamente al compito di eliminare l'ambiguità derivante da classi molto confuse. Si noti che tali classificatori possono operare sulla stessa finestra di elaborazione del modulo precedente, ma anche su opportune finestre, decise dal modulo centrale, per enfatizzare le parti del pattern dove si localizzano verosimilmente le differenze. È anche interessante notare che il processo di segmentazione dei caratteri è raffinato dagli autoassociatori che posizionano la finestra in un intorno della posizione indicata

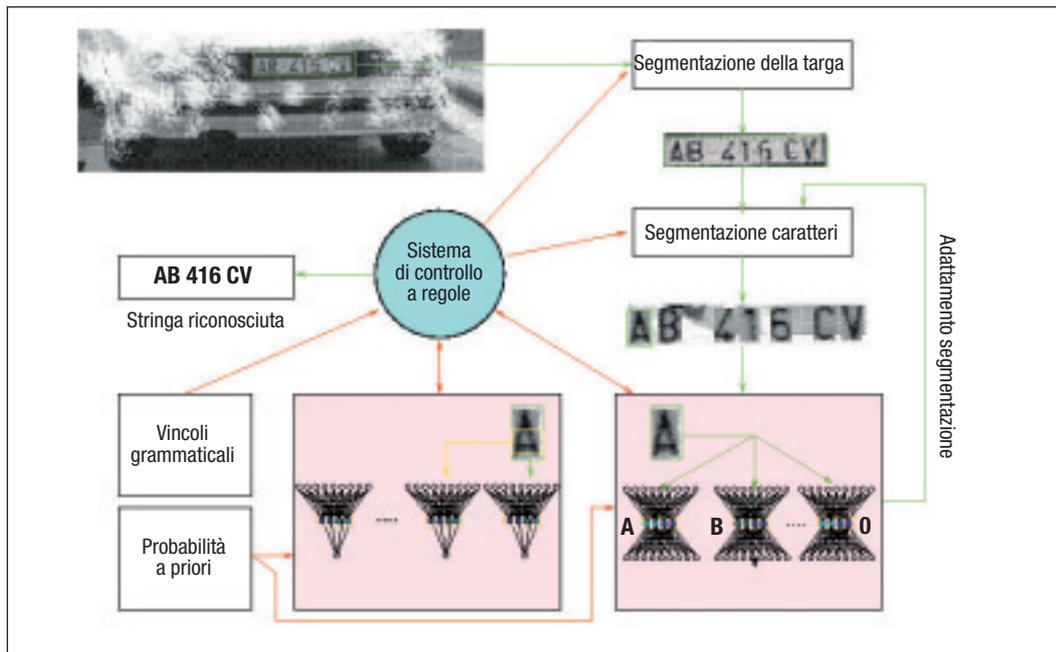


FIGURA 18
Architettura del sistema per il riconoscimento di targhe

dal modulo di segmentazione, a seguito di elaborazioni atte a stabilire la zona di massima risonanza.

7. LIMITI E PROSPETTIVE DI RICERCA

La ricerca nel settore delle reti neurali artificiali ha raggiunto un certo grado di maturità sia per quanto riguarda lo sviluppo sistematico delle metodologie fondamentali che il loro utilizzo in ambito applicativo. La comprensione dei limiti fondamentali sembra essere un passo fondamentale per lo sviluppo ulteriore del settore. Soprattutto in ambito applicativo, tali tecnologie sono state utilizzate talvolta in modo acritico confidando sul principio che la "forza bruta" derivante dall'impressionante sviluppo della microelettronica e il sogno della computazione inerentemente parallela potessero coniugarsi con l'apprendimento automatico per risolvere importanti problemi aperti con significativo risvolto applicativo. Un'analisi teorica dettagliata suggerisce la presenza di enormi ostacoli per un ulteriore sviluppo di approcci basati sugli attuali schemi di apprendimento automatico. Alcuni interessanti limiti erano già stati segnalati da Marvin Minsky nella sua edizione espansa di *Perceptrons* [10]. In particolare, Minsky aveva già individuato problemi legati

all'euristica del gradiente, alla base di molti schemi di ottimizzazione utilizzati per le reti neurali e dettati dall'esigenza di ottimizzare in spazi di enorme dimensione. Una volta formulato nell'ambito dell'apprendimento neurale, la complessità inerente di un problema si rivela in termini della dimensione dello spazio dei pesi e della forma della superficie errore da ottimizzare. L'ostacolo fondamentale per l'euristica del gradiente è dovuto alla presenza di minimi locali sub-ottimi che intrappolano gli algoritmi di apprendimento. Per problemi complessi, l'esplosione del numero di tali minimi locali rende inverosimile la determinazione di soluzioni efficienti. Si è già accumulata evidenza teorica e sperimentale che, a fronte di problemi "complessi", la forma della superficie errore si "regolarizza" e diminuisce la presenza di minimi sub-ottimi all'aumentare della dimensione dello spazio dei pesi. Tuttavia, tale aumento di dimensione non solo conduce a un incremento di complessità nel calcolo del gradiente, ma introduce il problema addizionale dell'*over-training*, secondo cui l'apprendimento in presenza di spazi dei parametri troppo grossi non garantisce una corretta generalizzazione a nuovi esempi. È forse il momento di costruire teorie computazionali dell'apprendimento adatte al calcolo neurale, tipicamente definito nel continuo, invece, che nel tradi-

zionale contesto discreto. Il PAC (*Probably Approximately Correct*) learning, ampiamente utilizzato fin qui per la comprensione della complessità, sembra piuttosto sterile e non appare molto efficace per la comprensione di tipici contesti applicativi⁶. Sembra, inoltre, importante procedere nella direzione di sviluppare architetture e algoritmi di apprendimento nel contesto di ingressi strutturati, opportunamente rappresentati. Questo favorisce per altro lo sviluppo di integrazioni più forti tra modelli simbolici e sotto-simbolici e sembra suggerire, in generale, la formulazione di teorie più generali per il trattamento di dati continui. A tal proposito, Margaret Boden, con riferimento alla novella del "Mago di Oz" scrive "[...] *the pretty creature was visibly the same horse, changing colour as it trotted along. ... AI is one beast, like the Wizard's pony*", [4] proponendo il parallelo dei colori del pony con i diversi colori dell'intelligenza artificiale. Servono forse nuove interessanti miscele di colori, servono schemi per modellare in modo più naturale l'incertezza, serve comprendere più a fondo l'ingrediente evolucionistico delle specie per coniugarlo con l'apprendimento automatico. E anche gli schemi di apprendimento devono verosimilmente risultare meno rigidi e, soprattutto, devono risultare attivi, permettendo un'interazione tra la macchina che apprende e il suo supervisore⁷. Forse non importa il "colore" del pony; le reti neurali, e più in generale la *computational intelligence*, devono integrarsi in modo più forte con i classici modelli simbolici. L'ibrido può non solo risultare vincente nelle applicazioni, ma può originare nuove miscele di colori, ben distinte dai componenti.

Bibliografia

[1] Ablameyko S., Goras L., Gori M., Piuri V.: *Limitations and Future Trends in Neural Computation*. IOS Publishing, (Eds 2003).

- [2] Anderson J., Rosenfeld E.: *Neurocomputing: Foundations of Research*. MIT Press, Cambridge, (Eds. 1988).
- [3] Angluin D., Smith C.: Inductive inference: Theory and methods. *Computing Surveys*, Vol. 15, n. 3, 1983, p. 237-269.
- [4] Boden M.: *Horses of a different colour? In Artificial Intelligence and Neural Networks*. V. Honavar and L. Uhr, Eds. Academic Press, 1994, p. 3-19.
- [5] Frasconi P., Gori M., Sperduti A.: A general framework for adaptive processing of data structures. *IEEE Transactions on Neural Networks*, Vol. 9, 1998, p. 768-786.
- [6] Hebb D.: *The Organization of Behavior*. Wiley, New York, 1949. Partially reprinted in [Anderson and Rosenfeld, 1988].
- [7] Hopfield J.: Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, Vol. 79, 1982, p. 2554-2558. Also in *Neurocomputing*, The MIT Press, 1988.
- [8] McCulloch W., Pitts W.: A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, Vol. 5, 1943. Reprinted in [Anderson and Rosenfeld, 1988].
- [9] Mead C.: *Analog VLSI and Neural Systems*. Addison Wesley, Reading, 1989.
- [10] Minsky M., Papert S.: *Perceptrons - Expanded Edition*. MIT Press, Cambridge, 1988.
- [11] Rosenblatt F.: *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanism*. Spartan Books, Washington D.C., 1962.
- [12] Rumelhart D., Hinton G., Williams R.: *Learning internal representations by error propagation*. In *Parallel Distributed Processing*, D. Rumelhart and J. McClelland, Eds. Vol. 1. MIT Press, Cambridge, Chapter, Vol. 8, 1986, p. 318-362. Reprinted in [Anderson and Rosenfeld, 1988].
- [13] Seeley R., Stephens T., Tate P.: *Essentials of Anatomy and Physiology*, McGraw-Hill, 2002.
- [14] Widrow B., Hoff M.: Adaptive switching circuits. In *IRE WESCON Convention Record*. IRE, New York, Vol. 4, 1960, p. 96-104.

MARCO GORI è professore ordinario all'Università di Siena presso il Dipartimento di Ingegneria dell'Informazione. Ha ottenuto il Dottorato di ricerca all'Università di Bologna, completando la formazione presso la "School of Computer Science" di McGill University, Montreal. I suoi interessi di ricerca riguardano l'intelligenza artificiale e le sue applicazioni. È attualmente chair del capitolo italiano della Neural Networks Society ed è presidente dell'Associazione Italiana Intelligenza Artificiale. marco@dii.unisi.it

⁶ Per lo stato dell'arte sui limiti e sulle prospettive del calcolo neurale si può far riferimento a Ablameyko *et al.* [1].

⁷ La ricerca nel settore del *learning from queries and examples* ha già fornito promettenti indicazioni di riduzione di complessità, ma sembra ancora essere in una fase incipiente, soprattutto per le ricadute applicative.