
Reti neurali come strumento di analisi delle facoltà cognitive: il problema della memorizzazione

Barbara Giolito
Università Vita-Salute San Raffaele, Milano
E-mail: barbara_giolito@libero.it



ABSTRACT: Uno dei principali e ricorrenti problemi di qualunque studio che voglia proporsi come un'analisi di carattere scientifico delle facoltà cognitive sembra consistere nella difficoltà a individuare strumenti che godano del rigore e della precisione imposti dalla scienza e, nello stesso tempo, risultino applicabili a facoltà psicologiche complesse quali quelle umane.

PAROLE CHIAVE: Neuroscienze, scienze cognitive.

1. Introduzione

Uno dei principali e ricorrenti problemi di qualunque studio che voglia proporsi come un'analisi di carattere scientifico delle facoltà cognitive sembra consistere nella difficoltà a individuare strumenti che godano del rigore e della precisione imposti dalla scienza e, nello stesso tempo, risultino applicabili a facoltà psicologiche complesse quali quelle umane. Per questa ragione il ricorso alle reti neurali in ambito psicologico e nel settore della filosofia della mente ha destato notevoli interessi: essendo le reti neurali modelli implementati attraverso strumenti informatici, esse richiedono infatti quella precisione ed esattezza che caratterizza la scienza e che spesso sembra rimanere un miraggio nello studio degli esseri umani in quanto soggetti mentali. Il ricorso alle reti neurali nello studio della mente ha tuttavia sollevato anche notevoli perplessità. Una critica che si sente spesso rivolgere ai modelli cognitivi che fanno uso di reti neurali consiste, in particolare, nel far notare come queste non siano in grado di riprodurre fedelmente le capacità mnemoniche degli esseri umani. La mente umana tiene traccia attraverso la memoria di molti degli aspetti relativi agli eventi con i quali, e all'interno dei quali, si trova a interagire. La memoria rappresenta, d'altra parte, un requisito basilare di molte delle capacità cognitive degli esseri umani: categorizzazione e linguaggio non potrebbero, ad esempio, essere realizzati senza di essa. Qualunque sistema incapace di dare luogo a prestazioni mnemoniche sufficientemente articolate non sembrerebbe pertanto poter rappresentare un buon candidato per lo studio delle funzioni mentali umane. Il ricorso a modelli a rete neurale è così stato, talvolta, disapprovato proprio sulla base dell'opinione secondo la quale abilità mnemoniche assimilabili a quelle umane non potrebbero essere implementate attraverso questi modelli: un caso limite in relazione ai problemi legati alle capacità mnemoniche delle reti neurali è il cosiddetto *catastrophic forgetting*, ovvero il fenomeno in base al quale le reti neurali non sarebbero soggette a un processo di oblio paragonabile al lento processo di oblio che caratterizza la vita mentale umana, ma darebbero luogo a una perdita dei dati memorizzati implausibilmente più veloce di quella che avviene negli esseri umani [5].

2. Reti neurali e memoria

Le reti neurali sono sistemi di calcolo distribuiti, le cui parti – costituite da *unità* e *connessioni* – operano in parallelo: ogni unità è un dispositivo di calcolo poco potente, che riceve in *input* valori numerici, li elabora e attraverso le connessioni li invia quali *output* alle altre unità. La potenza computazionale delle reti deriva dall'elevato numero di unità e connessioni che lavorano contemporaneamente. Le connessioni svolgono un ruolo attivo nei calcoli eseguiti dalle reti, in quanto modificano i dati trasmessi moltiplicandoli per i valori numerici (*pesi* o *forza* delle connessioni) ad esse associati. Altra caratteristica rilevante delle reti neurali è il fatto che in esse immagazzinamento ed elaborazione dei dati non siano demandati a meccanismi distinti: le unità rappresentano sia uno strumento di memoria sia uno strumento di calcolo. Inoltre, le reti sono in grado di *apprendere* nuovi compiti: i pesi delle connessioni sono generalmente lasciati evolvere durante la fase di calcolo in base a opportune regole – le cosiddette *regole di apprendimento*, che portano il risultato del calcolo ad avvicinarsi sempre di più ai dati desiderati – e questo rende possibile un fenomeno di modifica del comportamento delle reti paragonabile a una sorta di apprendimento [7][9][10]. L'utilizzo delle reti neurali come strumento di calcolo nello studio dei processi cognitivi presenta alcuni vantaggi: innanzitutto, la possibilità di generalizzare i dati ottenuti, estraendo da essi le informazioni ricorrenti per formare una sorta di *prototipo* degli

stessi, e la possibilità di trattare dati incompleti o parzialmente erronei (il grado di attivazione di un insieme di unità dipende dal contributo di più parametri e può quindi raggiungere livelli differenti a seconda di quanto suggerito globalmente da tali parametri, tentando una soluzione del problema posto alla rete anche quando alcuni dei valori da calcolare sono assenti o ingannevoli). Le reti neurali possono inoltre realizzare forme di rappresentazione distribuite, in cui ogni rappresentazione corrisponde a un *pattern di attivazione* ripartito su più unità: questa possibilità permette di lavorare a un livello che potremmo definire *pre-simbolico* o *pre-concettuale*, in cui ogni singola unità di elaborazione rappresenta, non il corrispettivo di un simbolo dotato di un proprio significato determinato, ma uno tra i suoi molteplici costituenti [1][4].

Abbiamo anticipato poco sopra come nelle reti neurali memoria e dispositivi di calcolo non siano separati e realizzati in componenti differenti ma immagazzinati nei pesi delle connessioni che collegano le unità. Consideriamo, ad esempio, il compito consistente nel decidere se l'input appena ricevuto – che possiamo immaginare come la rappresentazione di una certa forma geometrica – appartenga a una determinata categoria: l'esecuzione di tale compito può essere realizzata riconducendo il nuovo input a quello che è stato identificato come il prototipo della categoria geometrica in questione. La rete, in questo caso, non fa altro che manipolare il nuovo input attraverso le procedure codificate dalle sue connessioni. Poiché la memorizzazione dei dati precedentemente analizzati e relativi alla forma prototipica è stata realizzata nella rete attraverso la modifica dei pesi delle sue connessioni, i problemi sorgono, tuttavia, quando la rete deve apprendere un nuovo compito. Il nuovo compito viene, infatti, codificato – come accadeva per il precedente – attraverso i pesi delle connessioni tra le unità della rete: se i pesi vengono modificati per codificare la nuova procedura, essi non corrisponderanno presumibilmente più ai pesi che codificavano la procedura precedente. Dal momento che la memoria della rete era immagazzinata in questi ultimi, sembrerebbe lecito concludere che la stessa non possa che essere andata persa. Detto altrimenti, il fatto che le reti non possiedano veri e propri magazzini per la memoria ma la implementino sull'unità operativa sembra comportare, al sopraggiungere di un nuovo compito, una sovrapposizione delle nuove rappresentazioni mnemoniche sulle precedenti, sovrapposizione che porta all'eliminazione di queste ultime. Un simile fenomeno può costituire un problema nell'utilizzo delle reti neurali per lo studio delle funzioni cognitive umane: come abbiamo già detto, la memoria rappresenta non soltanto una funzione cognitiva tipica della mente, ma anche un presupposto indispensabile di molte delle facoltà mentali tipicamente umane. La memoria propria degli esseri umani è, d'altra parte, caratterizzata dal fatto di consentire un immagazzinamento complesso e multiforme dei dati relativi agli eventi incontrati nel passato: i nuovi dati di cui occorre tener traccia vanno in molti casi ad aggiungersi ai precedenti senza che questi ultimi siano eliminati. Apprendiamo quotidianamente nuovi compiti, che comportano la memorizzazione di nuovi elementi, senza che questo ci porti a dimenticare la corretta esecuzione dei compiti appresi nel passato.

3. Variazioni mnemoniche negli esseri umani: un paragone con le reti neurali

Per quanto una certa stabilità nelle capacità mnemoniche umane sia un fatto indiscutibile, non si può tuttavia sostenere che una vera e propria immutabilità dei dati mnemonici relativi alle esperienze passate, al sopraggiungere di nuove esperienze, sia un fenomeno assoluto. Innanzitutto, è sufficiente una breve riflessione per riconoscere come, in almeno alcuni casi per ognuno di noi, le più recenti esperienze possano avere modificato, spesso in modo inconscio, se non il contenuto dei ricordi passati, quantomeno il nostro modo di

guardare allo stesso e quindi il suo significato e la sua generale *atmosfera*. La conclusione positiva di un lungo esame getterà, ad esempio, con ogni probabilità un'aurea di positività sui ricordi – anche neutri – legati allo stesso, così come la delusione arrecataci da un amico tenderà a sbiadire la bontà delle passate esperienze comuni. Sembra pertanto che i ricordi non siano completamente impermeabili alle esperienze più recenti ma interagiscano con esse: i dati recentemente depositati nella nostra memoria sembrano scalfire la rigidità dell'immagazzinamento dei dati meno recenti. D'altra parte, negli esseri umani, la memorizzazione di nuovi elementi sembra poter comportare, non solo la limitata modifica, ma la sostituzione del prodotto di una passata memorizzazione. Per renderci conto di come questo possa avvenire, è opportuno guardare alla psicoanalisi: uno dei meccanismi in base ai quali opera il processo psicoanalitico è costituito dal tentativo di rimuovere il collegamento tra determinati eventi e le sfumature emotive negative e traumatizzanti che sono state a essi collegate. Un principio fondamentale della cura psicoanalitica del disagio psicologico consiste nel presupposto che quest'ultimo derivi dall'aver rimosso un'esperienza traumatizzante dalla nostra memoria consapevole per relegarla nella parte inconscia della vita mentale: tale esperienza non smette, in questo modo, di influenzare il nostro vissuto interiore, andando al contrario a ledere in modo ancora più significativo il nostro equilibrio proprio perché operante a un livello che, in quanto inconscio, non permette alcuna difesa da parte del nostro pensiero razionale. Portando il soggetto a rivivere interiormente l'evento traumatizzante e a prenderne coscienza, il processo psicoanalitico si basa sulla possibilità di sostituire una colorazione neutra alla caratterizzazione negativa di ciò che viene inconsapevolmente ricondotto al vissuto emotivo destabilizzante: in un certo senso, si potrebbe dire che il processo psicoanalitico costituisce un tentativo di *riprogrammare* una parte della nostra memoria inconscia. La memorizzazione di un collegamento non consapevole tra due eventi, uno dei quali caratterizzato da una connotazione altamente negativa, può infatti comportare che anche il ricorrere dell'evento privo di tale connotazione richiami a livello emotivo l'angoscia legata all'evento traumatizzante: una via per eliminare lo scatenarsi di tale angoscia può, pertanto, consistere nel sostituire la memorizzazione del collegamento tra i due eventi con la memorizzazione della loro effettiva separazione. Conducendo il soggetto a prendere coscienza del fatto che un certo fenomeno non è di per sé caratterizzato da alcuna connotazione negativa ma appare tale in quanto inconsapevolmente collegato a un evento traumatizzante, si può liberare il vissuto di tale evento dalle sfumature negative che erano state ingiustificatamente connesse allo stesso.

Un suggerimento verso l'interpretazione del procedimento psicoanalitico qui proposta sembra venire da Parisi [6]: questi ipotizza di guardare alle parole che lo psicoterapeuta rivolge al proprio paziente durante la seduta psicoterapica come a una serie di stimoli, paragonabili all'input di una rete neurale, i quali vanno a influenzare e modificare il modo di operare del sistema nervoso del paziente, paragonabile – a sua volta – al modo di procedere della rete. Le parole dello psicoterapeuta e la loro elaborazione da parte del paziente possono generare secondo Parisi un processo di mutazione, nel sistema nervoso del paziente, simile alle mutazioni che la manipolazione di un determinato input può causare nei pesi delle connessioni di una rete: questo tipo di mutazione è tale da produrre effetti specifici e duraturi ma richiede molto tempo. L'analogia tra l'influenza della psicoanalisi sul sistema nervoso e il funzionamento delle reti neurali chiarirebbe, peraltro, la possibilità di rendere conto del fatto che azioni tanto diverse quali le parole di uno psicoterapeuta e l'utilizzo di psicofarmaci possano lavorare assieme con le stesse finalità: le parole dello psicoterapeuta – paragonate all'input di una rete –

costituirebbero infatti, così come gli psicofarmaci, uno stimolo fisico elaborato dal sistema nervoso del paziente, con la differenza che gli psicofarmaci – che agiscono in modo più diffuso e meno specifico sul sistema nervoso umano producendo effetti rapidi ma meno puntuali – dovrebbero essere equiparati, non a uno specifico input, ma ad azioni operanti sullo stato globale della rete, quali un generale innalzamento o abbassamento delle soglie di attivazione delle sue unità.

Torniamo, allora, alle reti neurali e al fenomeno della perdita della memoria: se, come sembra suggerire la psicoanalisi, anche la mente umana è caratterizzata dal fatto che esperienze successive possono sostituire la memorizzazione delle precedenti, la presenza di un simile fenomeno nei modelli a rete neurale potrebbe non apparire come uno svantaggio per il loro utilizzo nello studio della mente ma, al contrario, come un loro vantaggio. A chi volesse rifiutare il ricorso alle riflessioni sul metodo psicoanalitico per sostenere un modello della mente che, come le reti neurali, prende come spunto la struttura del sistema nervoso potrebbe – d'altra parte – essere fatto notare come alcune delle più recenti riflessioni sulle basi neurofisiologiche del metodo psicoanalitico abbiano messo in luce la possibilità di trovare correlati neurali delle strutture mentali presupposte dall'applicazione di tale metodo [11]. I neuroscienziati hanno, ad esempio, identificato sistemi di memoria inconscia che sembrano giustificare reazioni emozionali apparentemente prive di spiegazione. Al di sotto della corteccia cosciente si trova un collegamento neuronale tra le strutture deputate all'informazione percettiva e quelle – primitive – responsabili delle reazioni di paura: poiché questo collegamento evita l'ippocampo, ovvero la struttura che genera i ricordi coscienti, esso può essere considerato come il responsabile dell'attivazione di *ricordi* di avvenimenti passati emotivamente significativi ma inaccessibili alla nostra consapevolezza. Le ricerche neuroscientifiche hanno, inoltre, dimostrato che le strutture cerebrali che supportano la formazione di ricordi coscienti non operano durante i primi due anni di vita degli esseri umani: un simile ritardo potrebbe essere considerato come il correlato neurale di quella che viene definita in psicoanalisi "amnesia infantile", ovvero l'incapacità di richiamare alla coscienza i ricordi della prima infanzia, i quali influenzano tuttavia la vita mentale adulta. Una conferma al fenomeno della rimozione – secondo il quale ricordi di cui non siamo consapevoli influiscono sulla nostra vita mentale – viene, infine, da alcuni esperimenti su pazienti anosognosici: a causa di un danno alla regione parietale destra del cervello, questi pazienti risultano essere inconsapevoli di handicap fisici anche gravi al lato sinistro del loro corpo. La stimolazione artificiale dell'emisfero destro riconduce tali pazienti alla consapevolezza, anche se momentanea, del loro handicap e di come questo si sia prodotto: questo fenomeno sembra rafforzare l'ipotesi che dati inaccessibili al pensiero consapevole vengano registrati e mantenuti a un livello inconscio.

4. Conclusione: reti neurali con adeguate capacità di memoria e oblio

Quanto detto finora sembrerebbe andare a sostegno dell'idea che il fenomeno di sostituzione della passata memorizzazione con quella più recente – fenomeno che contraddistingue il funzionamento delle reti neurali – caratterizzi anche parte della vita mentale umana, quella parte della quale si occupa la psicoanalisi, precedente alla riflessione razionale e che sembrerebbe trovare una giustificazione a livello neuronale in meccanismi di memorizzazione differenti dai sistemi correlati all'esperienza cosciente. La capacità di conservare i precedenti ricordi al sopraggiungere dei nuovi è, tuttavia, un'abilità innegabile degli esseri umani: essendo tale capacità un aspetto fondamentale della vita mentale, un modello incapace di riprodurlo non sembrerebbe poter rappresentare

uno strumento adeguato per l'analisi delle facoltà cognitive. A questo proposito, interessante è il fatto che recenti studi sul fenomeno del *catastrophic forgetting* abbiano mostrato come anche le reti neurali possano essere dotate di meccanismi tali da permettere loro una più plausibile conservazione dei dati memorizzati [3]. French è, ad esempio, partito dall'ipotesi che il fenomeno del *catastrophic forgetting* sia dovuto a un'eccessiva sovrapposizione delle rappresentazioni interne delle reti neurali: il problema sorgerebbe dalla natura pienamente distribuita delle rappresentazioni interne delle reti e potrebbe pertanto essere, almeno in parte, risolto attraverso algoritmi in grado di produrre rappresentazioni interne semi-distribuite. La soluzione proposta da French ricorre alla cosiddetta *tecnica degli pseudopattern* [8]: supponiamo che una rete con n input e m output abbia appreso a elaborare un insieme di pattern di input-output $\{P_1, P_2, \dots, P_N\}$ per mezzo di una funzione f ma che i vettori originari non siano più disponibili. Creiamo ora un numero M di vettori di input casuali di lunghezza N $\{i_1, \dots, i_M\}$ e li immettiamo nella rete: essa produrrà un insieme di output $\{o_1, \dots, o_M\}$ per ognuno degli pseudoinput. In questo modo potremo ottenere un insieme di pseudopattern $S = \{\psi_1, \psi_2, \dots, \psi_M\}$, dove $\psi_1: i_1 \rightarrow o_1, \psi_2: i_2 \rightarrow o_2, \dots, \psi_M: i_M \rightarrow o_M$, il quale dovrebbe approssimativamente corrispondere alla funzione originariamente appresa dalla rete. Quando la rete dovrà apprendere nuovi insiemi di pattern, un certo numero di questi pseudopattern potrà essere aggiunto ai nuovi pattern, in modo da impedire alla rete la perdita delle funzioni precedentemente apprese. A questo è possibile aggiungere la tecnica dei *modelli di memoria duale* [2]: questi modelli comprendono due aree di elaborazione dei pattern, separate ma continuamente interagenti l'una con l'altra. Un'area è utilizzata per l'elaborazione iniziale, mentre l'altra costituisce una sorta di magazzino a lungo termine: poiché l'informazione passa dall'una all'altra di tali aree attraverso gli pseudopattern precedentemente descritti, le reti così realizzate sono in grado di attuare un tipo di memoria dotato di oblio graduale. Le reti neurali, al contrario di quanto temuto, sembrerebbero pertanto capaci di riprodurre sia i meccanismi di *riprogrammazione* non consapevole propri del nostro inconscio sia le capacità mnemoniche che caratterizzano la nostra vita mentale consapevole.

Riferimenti bibliografici

- [1] E.A. Bates – J.L. Elman (1993), *Connectionism and the study of change*, in N.H. Johnson, *Brain development and cognition*, Blackwell, Oxford.
- [2] R.M. French (1997), *Pseudorecurrent connectionist networks: an approach to the "sensitivity-ability" dilemma*, in "Connection Science", 9;
- [3] R.M. French (2003), *Catastrophic forgetting in connectionist networks*, in L. Nadel, *Encyclopedia of Cognitive Sciences*, vol.1, Nature Publishing Group, London.
- [4] G.E. Hinton – J.L. McClelland – D.E. Rumelhart (1986), *Distributed representations*, in D.E. Rumelhart – J.L. McClelland, *Parallel Distributed Processing. Exploration the Microstructure of Cognition. Foundations*, vol.1, The MIT Press, Cambridge; trad. it. *PDP. Microstruttura dei processi cognitivi*, Il Mulino, Bologna, 1991.
- [5] M. McCloskey – N. Cohen (1989), *Catastrophic interference in connectionist networks: the sequential learning problem*, in G.H. Bower, *The psychology of learning and motivation*, vol.24, Academic Press, New York.
- [6] D. Parisi (2000), *La naturalizzazione degli esseri umani*, in "Sistemi Intelligenti", 1.

- [7] E. Pessa (1993), *Reti neurali e processi cognitivi*, Di Renzo, Roma.
- [8] Robins (1995), *Catastrophic forgetting, rehearsal, and pseudorehearsal*, in “Connection Science”, 7.
- [9] D.E. Rumelhart – J.L. McClelland (1986), *Parallel Distributed Processing. Exploration the Microstructure of Cognition. Foundations*, vol.1, *Psychological and Biological Models*, vol.2, The MIT Press, Cambridge; trad. it. *PDP. Microstruttura dei processi cognitivi*, Il Mulino, Bologna, 1991.
- [10] P. Smolensky (1988), *On the proper treatment of connectionism*, in “Behavioral and Brain Sciences”, 11; trad. it. *Il connessionismo tra simboli e neuroni*, Marietti, Genova, 1992.
- [11] M. Solms – O. Turnbull (2002), *The brain and the inner world*, Other Press, LLC, New York; trad. it. *Il cervello e il mondo interno*, Raffaello Cortina, Milano, 2004.