

Editoriale

Caro lettore,

questo numero, l'ultimo del 2018, prosegue la tradizione inaugurata ormai da qualche anno di un numero speciale, in cui l'informatica costituisce lo spunto per considerazioni culturali ad ampio raggio.

In questo numero ci concentriamo, ancora una volta, su un tema generale molto dibattuto, ossia la cosiddetta 'intelligenza' delle macchine, ma lo facciamo da una prospettiva specifica. Il tema dell'articolo principale è infatti il calcolo neuromorfico (*neuromorphic computing*), e in particolare l'approccio alla progettazione di circuiti elettronici che si ispirano al modo in cui il cervello umano elabora informazioni.

Il ruolo di ispirazione giocato dall'essere umano, e dal mondo biologico in generale, non è certo nuovo nella tradizione di ricerca che ha preso avvio con la cibernetica, prima, e con il modello basato su reti neurali artificiali nell'intelligenza artificiale poco dopo. Le incredibili capacità cognitive degli esseri umani, e la peculiarità della loro architettura cerebrale, hanno sempre rappresentato una fonte di ispirazione nella progettazione di artefatti con l'obiettivo di raggiungere prestazioni simili a quelle umane, almeno in certi specifici ambiti di applicazione.

L'articolo di Daniele Ielmini che pubblichiamo presenta in maniera chiara, e speriamo accessibile a tutti, lo stato di avanzamento della ricerca nell'ambito del calcolo neuromorfico, mostrandoci come le tecnologie future, i nuovi materiali e le nuove architetture dei processori stiano accelerando il passo delle innovazioni in questo affascinante settore tecnologico. Al contempo questi risultati sono presentati in un cornice storica che aiuta a meglio comprendere la strada che ha portato alla loro evoluzione.

Siamo alle soglie di una nuova intelligenza artificiale? I risultati che si stanno raggiungendo in questo settore consentiranno non solo enormi progressi tecnologici, ma anche di gettare luce sui meccanismi ancora ignoti della nostra cognizione? L'indagine sul rapporto mente-cervello che, per tanto tempo, è stata al centro del dibattito filosofico assumerà nuovamente importanza sotto questa diversa prospettiva?

Domande alle quali certamente non possiamo dare risposta, e che ci riconducono alla domanda sull'essenza del pensiero umano. Da un lato, il pensiero computazionale, basato su algoritmi, che certamente esercitiamo in gran parte

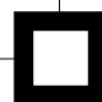
0

1

0

1

0



dei problemi che incontriamo ogni giorno. Dall'altro, il pensiero inventivo, creativo, libero che si interroga sul non spiegabile e scommette su una possibile interpretazione dell'ignoto.

Il nostro è dunque un invito alla lettura in cui ci auguriamo possiate trovare risposte rigorose alle domande che una delle sfide tecnologiche oggi in atto pone, ma anche spunti di riflessione sulle questioni più profonde che ci riguardano in quanto esseri umani: la conoscenza, il pensiero, la libertà. Anche per questo, a completamento del numero, vi proponiamo il racconto di una provocazione, molto nota nella storia dell'intelligenza artificiale, che ci invita a riflettere sulla differenza – se mai esiste – fra sembrare ed essere intelligenti.

Viola Schiaffonati

EDITORIALE

Roberto Saracco (EIT Digital)

A fine novembre 2017, dopo alcuni mesi di sperimentazione Verizon ha annunciato l'intenzione di iniziare il servizio commerciale 5G a metà 2018 in quattro aree degli US, a partire da Sacramento, CA. Due giorni dopo il CTO di T-Mobile ha dichiarato che l'annuncio di Verizon era un "wishful thinking" (proiezione di una speranza piuttosto che realtà) in quanto la tecnologia non è, né sarebbe stata pronta. Inoltre, ha aggiunto il manager, gli "economics" non tornano, al punto che T-Mobile, non ha alcuna intenzione di investire a breve nel 5G. Paradossalmente, hanno ragione entrambi! Verizon si propone di utilizzare lo spettro in campo millimetrico (28GHz) per portare la larga banda ad utilizzatori fissi disposti a mettere un ricevitore sulla finestra per catturare il segnale radio, che essendo in alta frequenza ha problemi a passare attraverso muri e vetri. Dal ricevitore verrebbe poi fatto passare un filo in casa con cui alimentare un hot spot WiFi. Non proprio quello che si immagina quando si parla di un sistema mobile.

ARTICOLI

5G Evoluzione o Rivoluzione?

Roberto Saracco (EIT Digital)

Il 5G è oggetto di una campagna mediatica che è più frutto di marketing che di evoluzione tecnologica, anche se questa esiste. Per il marketing il 5G è già qui, con Operatori che promettono offerte commerciali nel 2018, addirittura prima che gli standard vengano consolidati. Inoltre il marketing vende l'idea del 5G come presupposto alla evoluzione in settori come IoT, Industry 4.0 e self-driving cars. Ovviamente, l'industria non intende aspettare che il 5G si trasformi da marketing a realtà e sta muovendosi utilizzando le tecnologie attuali, compreso il 4G, che in massima parte soddisfa i requisiti richiesti dall'industria. Questo articolo vuole fare un reality-check mettendo anche in luce aspetti che il marketing trascura...

Il sistema 5G: una piattaforma progettata per i servizi digitali

Maurizio Décina (Politecnico di Milano)

Mai come per il 5G l'ecosistema industriale dei sistemi radiomobili cellulari si è mosso in forte anticipo e in sincronia verso la realizzazione di una sofisticata piattaforma multiservizio che permette la diffusione dei servizi Internet alle persone e alle cose. Un sistema guidato dalla più forte alleanza industriale del pianeta che, a meno di tre anni dalla sua nascita ufficiale (1-1-2020), incomincia a prendere forma concreta con i protocolli per l'IoT e le nuove interfacce radio. Gran parte dei servizi 5G potranno essere erogati con i sistemi 4,5G, del tipo LTE Advanced, ove i quattro principali ingredienti tecnologici della futura generazione 5G sono sperimentati e introdotti massivamente sul mercato: Piccole Celle, Virtualizzazione delle Funzioni di Rete, Multi-Access Edge Computing e Massive MIMO. Il 5G si differenzierà rispetto al 4,5G per le nuove interfacce radio NR (New Radio) e per i nuovi sistemi di codifica, multiplexing e correzione degli errori (Filtered OFDM, Sparse Code Multiple Access, Polar Codes, ecc.). L'obiettivo del 5G sono comunque i modelli di sviluppo per i servizi digitali offerti dagli operatori di telecomunicazioni in partnership con le utilities, le pubbliche amministrazioni, le imprese e gli Over The Top (OTT).

5G e fibra: dove e come investire?

Franco Bassanini (Open Fiber)

L'infrastruttura fissa, formata dalla rete in fibra, e quella mobile che abilita il 5G con una copertura pervasiva saranno cruciali nelle prossime due decadi per promuovere la crescita economica e consentire la fornitura di nuovi servizi a beneficio della clientela residenziale e business. Lo sviluppo di queste infrastrutture richiederà investimenti significativi, anno dopo anno. Al tempo stesso questi investimenti, se effettuati in modo corretto, si ripagheranno incrementando il PIL. Le stime su un campione di nazioni indicano che un incremento di investimento dell'1% del PIL si traduce in un incremento del PIL dello 0,4% nel primo anno e del 1,5% nei quattro anni successivi. Tuttavia, non sarà semplice attirare il capitale necessario alla costruzione delle infrastrutture in fibra e quelle per il 5G. In questo giocherà un ruolo importante il quadro regolatorio per stimolare iniziative di Public Private Partnership. Inoltre queste infrastrutture porteranno a cambiamenti dirompenti nel mercato delle infrastrutture di telecomunicazioni, rendendo ulteriormente complicata la transizione. Questo articolo affronta questi aspetti e propone un approccio integrato allo sviluppo della infrastruttura fissa e mobile.

Digital cities leveraging on 5G

Stéphane Péan (EIT Digital)

Due terzi della popolazione mondiale abiterà in contesti urbani nel 2030. Le città hanno progressivamente aumentato il loro peso economico, politico e sociale nel corso dei secoli e millenni e oggi sono all'avanguardia del mondo moderno, formando degli ecosistemi innovativi che generano prosperità. La trasformazione digitale è una delle principali evoluzioni odierne nello sviluppo delle città. Il suo obiettivo è l'ottimizzazione dei processi decisionali e l'empowerment della società civile. Nuove, creative innovazioni e nuovi modelli di business stanno emergendo ma spesso ad un ritmo troppo lento e non economicamente sostenibile. Occorre passare a volumi di scala, estendere quanto viene fatto ad altre città attraverso un riutilizzo efficace tramite piattaforme aperte e inserire il tutto in un tessuto di comunicazione ad alta capacità e basso costo. Il 5G all'orizzonte, parte della strategia di evoluzione europea, può essere un componente cruciale per stimolare la crescita dei servizi sia in termini di offerta che di domanda e può avere un impatto positivo sull'economia e sulla società urbana.

IEEE 5G and Beyond Initiative – A Perspective

Ashutosh Dutta (AT&T Labs), Gerhard Fettweis (TU Dresden)

Il 5G non rappresenta solo l'evoluzione del 4G ma un cambio di paradigma. Il 5G fornisce una maggiore capacità di banda e una più bassa latenza delle attuali generazioni ed ha una architettura in grado di abilitare nuove applicazioni soddisfacendo requisiti molto più stringenti consentendo l'esecuzione di funzionalità in tempo reale a larga banda. La 5G Initiative lanciata dalla FDC (Future Direction Committee dell'IEEE) ha un approccio a 360° sul tema andando oltre gli aspetti tecnologici per affrontare e promuovere la standardizzazione, lo sviluppo di tutorial e corsi di formazione, la creazione di una forte comunità di specialisti attraverso workshop, conferenze e uno spazio per la collaborazione on line che sta coinvolgendo quasi 10.000 membri.

5G beyond radio access: A flatter sliced network

David Soldani (Nokia)

Questo articolo descrive i principali scenari di utilizzo e i relativi requisiti tecnici per il 5G, presentando le diverse opzioni architetturali, tecnologie abilitanti, le bande spettrali definite dal 3GPP e le relative roadmap. Vengono affrontati sia gli aspetti radio (copertura, capacità e latenza) sia quelli della core network (slicing, architetture, protocolli). Viene illustrato in dettaglio la differenza tra LTE e 5G, sia a livello radio sia a livello rete core. Inoltre viene sottolineato come con il 5G il Total Cost of Ownership scenda sostanzialmente e come si creino nuove opportunità di ritorni, specialmente nel settore B2B, in particolare con offerte di Network Slice come Servizio (NSaaS) a prezzi competitivi per una varietà di settori.

EDITORIALE

Metodi e tecnologie per l'uso educativo e didattico dei robot

Edoardo Datteri, Luisa Zecca (Università degli Studi di Milano-Bicocca)

L'espressione "robotica educativa" indica l'utilizzo di robot come strumenti mediatori dell'apprendimento in contesti scolastici o extra-scolastici. Le attività generalmente proposte agli studenti comportano la realizzazione e la sperimentazione di programmi che permettono al robot di reagire in modi desiderati agli stimoli sensoriali, nonché, in alcuni casi, la costruzione fisica del robot stesso. Queste attività si basano sull'assunzione secondo cui la costruzione e la programmazione robotica, sotto certe condizioni, possano "allenare" capacità visuo-spaziali, ragionamento e creatività nelle loro molteplici dimensioni legate all'osservazione, previsione, scoperta e spiegazione dei fenomeni e all'identificazione ed esecuzione di strategie per risolvere problemi di vario genere. Molti insegnanti ritengono inoltre che la costruzione e la programmazione robotica possano facilitare lo sviluppo di capacità e competenze di natura disciplinare, soprattutto legate all'area scientifica, matematica e tecnologica. Sono convinzioni davvero fondate? Questa domanda viene affrontata, da vari punti di vista, negli articoli raccolti in questo numero speciale, che articolano alcuni contributi presentati al Convegno nazionale dal titolo "Giocare a pensare. Metodi e tecnologie per l'uso educativo e didattico dei robot" organizzato dall'Università degli Studi di Milano-Bicocca (20 Maggio 2017).

ARTICOLI

Robot attorno a noi: dove sono, cosa fanno, cosa faranno?

Andrea Bonarini (Politecnico di Milano)

Stiamo assistendo all'invasione silenziosa dei robot nella vita di tutti i giorni. Dall'originale impiego in fabbrica, i robot vedono ormai una crescita esponenziale al di fuori dell'ambito della produzione industriale. È importante quindi capire come fa un robot autonomo a svolgere il proprio compito, quali sono le possibilità reali di applicazione in questo momento, e quali potranno essere in un prossimo futuro. Presentiamo una rapida rassegna dello stato della robotica ad oggi, e degli sviluppi in un futuro prossimo. Infine, mostriamo alcune implicazioni sociali ed etiche che stanno emergendo di conseguenza.

Il progetto eCraft2Learn: Fabbricazione Digitale e Movimento dei Maker in ambito educativo

Francesca Agatolio, Emanuele Menegatti, Michele Moro (Università di Padova)

eCraft2Learn è un progetto europeo H2020 di due anni nato per promuovere l'interesse degli studenti di 13-17 anni verso le materie STEAM attraverso l'uso di tecnologie a basso costo, componenti elettroniche DIY e stampanti 3D. L'obiettivo principale del progetto è la progettazione, la creazione e la validazione di un ecosistema che integri queste tecnologie al fine di supportare la costruzione di artefatti digitali all'interno di una cornice metodologica basata sulla prospettiva costruzionista e ispirata al movimento dei maker. L'articolo contiene una descrizione delle premesse iniziali, del modello pedagogico adottato e alcuni dettagli riguardanti l'architettura dell'ecosistema.

Laboratori di robotica ed educazione tecnologica contestualizzata

Paola Mengoli e Margherita Russo (Università di Modena e Reggio Emilia)

Il saggio descrive i laboratori di robotica di Officina Emilia, realizzati tra il 2009 e il 2013 come ricerca-azione a sostegno del sistema di istruzione, di fronte alle sfide della nuova manifattura di industria 4.0. Vengono presentati metodi e risultati della valutazione delle attività e si conclude con una riflessione sulle modalità di coinvolgimento degli studenti, dei loro docenti e delle scuole, con l'obiettivo di creare contemporaneamente apprendimenti significativi, esperienze di formazione in servizio per i docenti, innovazione delle pratiche didattiche.

L'uso ludico-didattico dei robot in pediatria. Una ricerca esplorativa sul benessere dei bambini ricoverati

Marco Castiglioni, Giulia Zappa, Alessandro Pepe (Università degli Studi di Milano-Bicocca)

Scopo dello studio è valutare, mediante un piano d'indagine a matrice quali-quantitativa, l'impatto di un intervento ludico-didattico effettuato con il robot CoderBot presso un reparto di pediatria per degenze brevi. I risultati della ricerca, tuttora in corso, supportano l'idea che l'attività con il robot sortisca effetti positivi sul benessere emotivo dei partecipanti. Si rileva un effetto significativo dell'età sui punteggi rilevati prima e dopo l'attività: l'intervento appare «ottimale» per la fascia 4-10 anni. Dai dati qualitativi si evince che l'attività robotica ha come punti di forza novità e appeal tecnologico.

Introduzione della robotica in attività di problem solving nella scuola primaria. Analisi dei livelli di focalizzazione sugli scopi del problema

Andrea Garavaglia, Livia Petti, Emiliana Murgia, Francesca Bassi, Samantha Lisa Maranesi (Università degli Studi di Milano-Bicocca)

L'obiettivo della ricerca è quello di analizzare il primo utilizzo della robotica nella scuola primaria nell'ottica di individuare i livelli di focalizzazione sul problema principale proposto didatticamente rispetto alla necessità di dominare la complessità del sistema composto dalla novità di elementi tecnologici complessi quali i robot e la sua programmazione. A livello metodologico sono state previste l'osservazione delle attività, la somministrazione di prove di comprensione del problema e le interviste a un campione di bambini per individuare i livelli di attenzione sul problema principale.

I bambini e la robo-etologia: riflessioni epistemologiche sugli stili di spiegazione del comportamento dei robot

Giovanna Di Maio, Emiliana Murgia, Edoardo Datteri (Università degli Studi di Milano-Bicocca)

Le attività didattiche basate sull'uso di robot come mediatori dell'apprendimento coinvolgono spesso momenti di osservazione e spiegazione del comportamento del robot stesso. Quali sono i principali stili di spiegazione adottati da bambini di Scuola Primaria in questi contesti? In questo articolo si cercherà di affrontare questa domanda elaborando, anzitutto, una tassonomia di stili di spiegazione radicata nella letteratura epistemologica e distinta in alcuni sensi interessanti da altre tassonomie proposte nella letteratura sull'interazione bambino-robot. La tassonomia verrà poi esemplificata e messa alla prova nella codifica di un'esperienza in cui è stato chiesto a bambini di una classe prima di Scuola Primaria di descrivere e spiegare il comportamento di un robot pre-programmato.

Robotica e incremento di abilità visuo-spaziali in bambini con disabilità motoria nella scuola dell'infanzia

Valentina Pennazio (Università di Macerata)

Il contributo presenta gli esiti di sessioni sperimentali realizzate in alcune scuole dell'Infanzia della Liguria che hanno previsto l'impiego della robotica in attività didattiche di piccolo gruppo per ridurre le difficoltà visuo-spaziali tipiche dei bambini con grave disabilità motoria (conseguenti a forme severe di Paralisi Cerebrale Infantile) favorendone l'inclusione. Gli esiti delle sessioni sperimentali hanno dimostrato che un robot telecomandato mediante sensori, agendo come una sorta di prolungamento del corpo, può consentire ad un bambino con grave disabilità motoria di sviluppare una maggiore conoscenza dello spazio e abilità visuo-spaziali utili in vista di una successiva autonomia di movimento.

CoderBot: un robot didattico open source

Roberto Previtera (ICT project manager)

L'articolo descrive il progetto di ricerca che ha portato alla realizzazione e impiego di un robot didattico che potesse ovviare ai principali limiti dei prodotti esistenti, in particolare in termini di facilità di programmazione e flessibilità di utilizzo, e consentire attività didattiche nuove e stimolanti nel contesto di un progetto di sviluppo "open source" ovvero liberamente utilizzabile ed estendibile dalla comunità di utilizzatori: università, scuole, educatori. Il robot è stato impiegato in diverse sessioni didattiche con alunni della Scuola Primaria; si è nel tempo evoluto rispetto al progetto iniziale in modo da soddisfare i requisiti dei progetti didattici definiti da pedagogisti e insegnanti.

EDITORIALE

Il “ladro di numeri” restituisce saggezza digitale

Bruno Lamborghini

E' uscito un bel libro intitolato *Il ladro di numeri*, scritto da Norberto Patrignani, Mirella Mazzarini e Daniela Cappelletti, edito nell'ambito del Progetto formativo su Scuola, lavoro e territorio del Gruppo Loccioni di Jesi (www.loccioni.com), un gruppo che opera a livello internazionale nella progettazione di sistemi di automazione e controllo ed è una grande azienda-scuola-laboratorio che opera con valori olivettiani, ma soprattutto è fortemente coinvolta nella formazione manageriale e per l'alternanza scuola lavoro nello scenario digitale.

ARTICOLI

Apologia della ragione scientifica – III: decisioni e giochi strategici

Thomas Bassetti (Università degli Studi di Padova) e Angelo Luvison (AICA)

L'Homo oeconomicus – il risultato di un mero costrutto mentale – si trova costantemente di fronte a circostanze che ne sfidano la razionalità. A partire dal quadro concettuale fornito dalla teoria dei giochi, si illustrano alcuni dilemmi emblematici che si presentano in situazioni reali: economia, telecomunicazioni, società civile e politica. Il Leitmotiv dell'articolo è che l'esercizio del pensiero critico (critical thinking) e della razionalità, insieme agli strumenti disponibili (analisi quantitativa e probabilistica, rigore metodologico e logico, pragmatismo, ecc.), consente di inquadrare, affrontare e risolvere efficacemente i problemi che caratterizzano la trasformazione digitale, nel presente e nel futuro. Gli esempi a supporto della tesi sono di teoria dei giochi, paradossi e dilemmi, inclusi quelli economici, sociali ed elettorali.

Il computer fra fantasia e realtà

Franco Filippazzi (AICA)

Una macchina come il computer colpisce fortemente l'immaginazione e si presta alle estrapolazioni più avveniristiche. Non poteva quindi il computer non avere ampio spazio nella letteratura di fantascienza. In questo scritto si delineano alcuni tratti delle fantasie suscitate da questa macchina, fantasie che costituiscono non di rado spunti di riflessione e anche anticipazioni del futuro.

BEST PAPERS DIDAMATICA 2017

Per una didattica del latino tra conoscenze disciplinari e competenze digitali

Alessandro Iannella, Giuseppe Fiorentino, Isabella Pera Il contributo documenta un'esperienza di didattica del latino progettata come risposta allo scarso rilievo dato alla materia dalla nuova configurazione del liceo linguistico. La programmazione didattica, strutturata per specifiche competenze, ha avuto come obiettivo quello di fornire agli studenti non solo conoscenze disciplinari ma anche abilità digitali e metacognitive. Grazie all'impiego di un insieme di metodi e di strumenti in grado di offrire un approccio innovativo allo studio della materia, sono state realizzate attività di immersive education (VR, AR), peer e self assessment, gamification, digital storytelling e authentic learning, coordinate all'interno di una cornice pedagogica di carattere induttivo-contestuale. L'esperienza ha verificato la fattibilità di un percorso alternativo dimostrando il potenziale del blended learning in un'area disciplinare, tutto sommato, ancora lontana dall'innovazione.

Polireading: alzare il livello della discussione pedagogica

Aldo Torrebruno, Anna Torrebruno, Stella Casola

L'articolo illustra i primi risultati del progetto Polireading, in corso di realizzazione da HOC-LAB del Politecnico di Milano su sollecitazione del Cantiere Scuola Digitale di Forum PA. L'idea di fondo è quella di offrire ogni mese alcuni articoli scientifici tradotti in italiano su un determinato argomento stimolare la discussione, utilizzando gli strumenti del cloud computing, tra i docenti e i ricercatori. Dopo un mese, la discussione viene chiusa in scrittura e viene elaborato un riassunto “a più voci” della stessa. L'articolo descrive i risultati dei primi mesi, raccolti attraverso un questionario online ed evidenzia come Polireading sia un progetto che sta contribuendo ad innalzare in maniera significativa il dibattito sulle tecnologie per la didattica, colmando un bisogno nettamente percepito dai docenti italiani.

RUBRICHE

ANNIVERSARI

50 anni di Premio Turing: una riflessione

Daniel Russo

RACCONTI BREVI

Pasifæ Inc.

Alberto Cammozzo

In collaborazione con:



EDITORIALE

Didamatica 2018

G. Adorni, M. Cicognani, F. Koceva, G. Mastronardi

Promosso annualmente da AICA, DIDAMATICA (DIDAttica e inforMATICA – Informatica per la Didattica) da oltre trent'anni è punto di riferimento per studenti, docenti, istituzioni scolastiche, professionisti ICT, aziende e Pubblica Amministrazione sui temi dell'innovazione digitale per la filiera della formazione. La manifestazione è organizzata annualmente da AICA, in collaborazione con il MIUR. L'edizione 2018 si è svolta nel Campus di Cesena dell'Università di Bologna. Tema del convegno è "Nuovi Metodi e Saperi per formare all'Innovazione". La riflessione che DIDAMATICA 2018 intende sollecitare ha come scenario la società digitale e il mondo del lavoro, binomio che richiede un articolato ventaglio di competenze trasversali come, ad esempio, problem solving, pensiero laterale e capacità di apprendere. La crescita della società digitale non può che partire dall'istruzione e dalla formazione, ponendo al centro dei processi di apprendimento non solo la tecnologia, ma nuovi modelli di interazione allievo docente-realtà sociale.

SESSIONI

Prepararsi a vivere nel nuovo mondo digitale

Pensiero computazionale, Coding, Making e Robotica Educativa

I nuovi orizzonti tecnologici e metodologici per la formazione

Innovare la formazione per formare all'Innovazione

Nuove soluzioni formative di supporto all'alternanza scuola-lavoro

Poster



www.aicanet.it/didamatica2018

In collaborazione con:



EDITORIALE

Una storia da non dimenticare

Giancarlo Vaccari

L'ELEA 9003, matricola 2, oggi installata presso l'ITIS di Bibbiena, è l'unico esemplare, ancora esistente e in gran parte funzionante, di una serie di elaboratori elettronici sviluppati dall'Olivetti Divisione Elettronica negli anni 50/60 che, all'epoca, erano tra i più tecnologicamente avanzati. L'ELEA 9003 fu il primo elaboratore elettronico al mondo, di grandi dimensioni, costruito interamente con componenti a stato solido (transistor e diodi). Ha quindi un indubbio valore storico, a testimonianza dello spirito innovativo e della lungimiranza di Adriano Olivetti e del valore della squadra di ingegneri e tecnici guidati da Mario Tchou, che avevano sviluppato il progetto e portato sul mercato un prodotto al massimo delle prestazioni a livello mondiale.

ARTICOLI

Il teatro fra scienza e fantascienza

Giuseppe Longo (Università degli Studi di Trieste)

Dopo alcune considerazioni sul rapporto tra scienza e teatro, in cui si rilevano i pregi, ma anche i limiti e le difficoltà della comunicazione della scienza attraverso la scena, si prendono in esame due drammi: Minnie la candida di Massimo Bontempelli (1927) e R.U.R. di Karel Čapek (1920), analizzandoli e mettendone in luce i numerosi aspetti che anticipano temi, attuazioni e problemi attuali.

LEO, Olivetti e i primi computer commerciali in Europa: un'occasione perduta?

Elisabetta Mori (Middlesex University)

LEO è l'acronimo di Lyons Electronic Office, una serie di computer prodotti dalla nota compagnia britannica di catering J. Lyons & Co, fondata a Londra nel 1894. LEO I è stato riconosciuto come il primo business computer al mondo, operativo nel Regno Unito dalla fine del 1951, alcuni mesi prima del più celebre omologo americano UNIVAC I. L'articolo ripercorre la storia di LEO Computers dal 1947 - inizio della collaborazione con l'Università di Cambridge - fino al 1963, anno in cui LEO Computers fu coinvolta nella prima di una serie di fusioni con altre imprese, culminate nel 1968 nella nascita di International Computers Limited (ICL), incluso un tentativo di partnership tra imprese europee alla fine del 1962, che coinvolse anche English Electric, Olivetti, Bull e Siemens.

A fil di dato: valore e comunicazione dell'informazione al tempo dello Human-Data Interaction Design

Angela Locoro (Università Carlo Cattaneo - LUIC)

L'articolo propone una panoramica dell'uso dei dati basata sulla loro "catena del valore" e sull'importanza che tutti gli interessati a vario titolo al loro uso siano dotati di strumenti adatti a un'interazione diretta, sensata, fruttuosa e allineata ai propri scopi; a partire dai luoghi dove essi hanno origine, a quelli in cui avviene il loro trattamento formale, alla libera circolazione e disponibilità nei formati aperti del Web, il destino dei dati è quello di influenzare, con nuove forme di espressione accessibili a tutti, le nostre vite, la nostra conoscenza del mondo e le nostre decisioni più importanti.

RUBRICHE

STORIA DELL'INFORMATICA

La calcolatrice di Leonardo

Silvio Hénin

PROGETTI

Una passeggiata nell'archivio di *Il Cartastorie*

Olga Lo Presti

COMPUTER HUMOR

Quando le vignette ci raccontano il computer

Franco Filippazzi

In collaborazione con:



Intelligenza artificiale: l'approccio neuromorfico

Daniele Ielmini

Sommario

Il cervello umano è una macchina estremamente complessa, che, grazie alla sua architettura e alla modalità di calcolo, è in grado di risolvere problemi relativamente ardui in modo veloce e con un consumo di energia relativamente basso. Riconoscere un oggetto oppure un volto così come pilotare il nostro corpo in risposta ad uno stimolo sensoriale sono azioni tanto naturali per noi, quanto onerose dal punto di vista computazionale. Ricreare questo tipo di computazione veloce ed energeticamente efficiente è stato per decenni un obiettivo visionario della ricerca. Oggi, questo sogno si va concretizzando sempre più, grazie alla maturità delle tecnologie microelettroniche e ai progressi sul fronte delle reti neurali e neurobiologiche.

L'obiettivo di questo articolo è fare il punto della situazione sull'attuale stato in materia di calcolo neuromorfico. Si ripercorreranno le pietre miliari di questa disciplina e si passeranno in rassegna le principali tecnologie, sia in ambito software che nell'ambito della realizzazione di circuiti hardware in grado di emulare il modo in cui il cervello elabora le informazioni. Si accennerà infine alle sfide attuali, compresa quella di una approfondita conoscenza dei processi cognitivi nel cervello umano, e di come le tecnologie future, i nuovi materiali, e le nuove architetture potrebbero accelerare il passo delle innovazioni in questo affascinante settore tecnologico.

Abstract

The human brain is an extremely complex machine, which, thanks to its architecture and way of computation, is able to solve relatively complicated problems with high speed and relatively small energy consumption. Recognition of an object or a face, and control of our

body in response to a sensory stimulation, are as straightforward for us as complex from the computational viewpoint. Recreating this type of computation, which can be fast and energy efficient, has been a visionary objective of the research for decades. Today, this dream is coming true thanks to the maturity of the microelectronic technology and the progress in neural and neuro-biological networks.

The aim of this work is to review the state of the art of neuromorphic computing. The historical milestones of this topical area will be reviewed, and the main technologies, both in terms of software and in terms of hardware circuits that can mimic the way of reasoning of the brain, will be summarized. The future challenges will be described, including the need for a deeper understanding of the cognitive processes in the human brain, and the ability of future technologies, new materials, and new architectures, to accelerate the progress of the innovations in this fascinating technological field.

Keywords: deep learning; neuromorphic computing; spiking neural network; synaptic plasticity; memristor.

1. Introduzione

Il cervello umano è stato definito come la macchina più complessa dell'universo in quanto è noto a noi stessi ancora meno delle zone più remote del nostro pianeta. Frutto di milioni di anni di evoluzione naturale, il nostro cervello è in grado di riconoscere immagini, comprendere il linguaggio, controllare il movimento del corpo, prendere decisioni, pensare in termini astratti, ma soprattutto di apprendere dall'esperienza, abilità che lo contraddistingue e che lo rende intellettivamente superiore al cervello di ogni altra specie vivente. In particolare, ciò che più colpisce è che questa vasta gamma di funzioni è eseguita in un volume equivalente ad una scatola di scarpe, e con un consumo complessivo di potenza di circa 20 W, lo stesso di una lampadina ad incandescenza [1].

Sebbene sia essenzialmente uno strumento di elaborazione delle informazioni, il cervello differisce radicalmente dai computer che ci circondano. Fin dagli albori dell'era digitale, infatti, i computer furono progettati per elaborare informazioni binarie mediante operazioni Booleane. Inoltre, l'architettura dei computer convenzionali si basa sulla teoria di von Neumann, che prevede una profonda distinzione tra l'unità di elaborazione, dove vengono eseguite le operazioni, e la memoria, dove vengono immagazzinate le informazioni. Il cervello umano, al contrario, non è organizzato per compartimenti stagni, ma elabora l'informazione sensoriale in una rete biologica di neuroni collegati da connessioni sinaptiche. Ed è proprio questa architettura che assicura al nostro cervello l'elevata efficienza energetica che lo contraddistingue.

Lo sviluppo di una macchina in grado di replicare l'abilità computazionale del cervello è stato un costante obiettivo della ricerca ben prima dell'avvento dei

computer digitali. Tuttavia, è solo di recente che la tecnologia microelettronica, combinata ad una minima comprensione dei meccanismi di funzionamento dei neuroni e delle sinapsi, ha reso questo obiettivo realizzabile nella pratica. Oggigiorno, l'intelligenza artificiale (*artificial intelligence*, AI) sta entrando sempre più nelle nostre vite, abilitando operazioni relativamente semplici, come il riconoscimento facciale e vocale, ed in prospettiva sempre più pervasive, come l'uso di automobili e robot a guida autonoma. Nell'ambito dell'AI, l'ingegneria neuromorfica si propone di sviluppare circuiti di calcolo che imitino il cervello umano fin nei suoi aspetti più intimi, come il fenomeno di spike neuronale, l'architettura a rete neurale, e le regole di apprendimento mediante plasticità sinaptica. Tuttavia, nonostante i numerosi sforzi in questa direzione, un vero e proprio computer in grado di 'pensare' come il cervello umano è ancora di là da venire.

L'obiettivo di questo articolo è fare il punto della situazione sull'ingegneria neuromorfica da vari punti di vista: biologico, matematico, teorico, circuitale, e tecnologico. Verranno ripercorse le principali tappe dello sviluppo del calcolo neuromorfico, in termini di architetture, approcci progettuali, implementazioni hardware, e nuovi materiali che possono accelerare lo sviluppo di reti neuromorfiche ad alta densità. Le attuali sfide, le nuove tecnologie per superarle e i possibili scenari futuri verranno infine riassunti in uno sguardo d'insieme.

2. Breve storia del calcolo neuromorfico

La scienza ha da sempre cercato di comprendere e modellare la mente umana ed i relativi meccanismi fisici, chimici e biologici. Già alla fine del XIX secolo, Santiago Ramón y Cajal riuscì a formulare un'immagine precisa del neurone come mattone fondamentale del cervello grazie alla tecnica di colorazione precedentemente sviluppata da Camillo Golgi. Cajal, che per questi progressi fu insignito del premio Nobel per la medicina nel 1906, sfruttò la sua abilità di disegnatore per fornire una rappresentazione più realistica del neurone dove, per la prima volta, fu evidenziata la sua struttura ramificata (Fig. 1a), suggerendo quindi una complessa rete neurale di trasmissione ed elaborazione parallela delle informazioni.

Le scoperte di Cajal furono il punto di partenza per il primo modello matematico del neurone, sviluppato nel 1943 da Warren S. McCulloch e Walter H. Pitts (Fig. 1b) [2]. Nel loro modello, il neurone è descritto da una funzione matematica, la cui variabile indipendente è la somma dei segnali provenienti da vari neuroni di ingresso, mentre la variabile di uscita è un segnale fortemente non lineare, ad esempio una funzione di attivazione sigmoideale. Sebbene fortemente idealizzata, la funzione neuronale di McCulloch e Pitts evidenzia una forte analogia con la natura biologica del neurone. Modelli neuronal più sofisticati, come quello di Hodgkin e Huxley ricavato dallo studio degli assoni giganti dei calamari, sono in grado di descrivere sia la natura transitoria dell'impulso (*spike*) neuronale, sia i dettagli chimici legati al rilascio degli ioni calcio e potassio [3]. D'altro canto, il pregio del modello di McCulloch e Pitts era che ogni singolo neurone veniva considerato nel contesto dell'interazione con

altri neuroni, per descrivere non solo la struttura morfologica, ma soprattutto la funzione matematica di elaborazione del segnale all'interno del cervello.

Il maggiore progresso in questa direzione venne dal concetto di 'perceptrone', introdotto per la prima volta da Frank Rosenblatt nel 1957 [4]. Il perceptrone non è nient'altro che una semplice rete neurale, essenzialmente la stessa di Fig. 1b, in grado di operare una classificazione lineare dei dati in ingresso, come ad esempio le immagini. La novità distintiva è che Rosenblatt introdusse anche un algoritmo di apprendimento, che permette al perceptrone di adattare i pesi sinaptici al fine di separare due classi di oggetti, ad esempio automobili e motociclette (Fig. 1c). Il perceptrone è quindi un passo avanti fondamentale verso sistemi neuromorfici in grado di elaborare l'informazione come nel cervello umano, introducendo aspetti chiave come il parallelismo e la plasticità sinaptica, che sono caratteristiche peculiari delle reti neurali biologiche.

Il limite principale del perceptrone di Rosenblatt consiste nella possibilità di classificare con successo soltanto famiglie linearmente separabili, che possono cioè essere riconosciute sulla base di una semplice disequazione $y > f(x)$ dove $f(x)$ è una funzione lineare dei segnali neuronali x di ingresso (Fig. 1c). In presenza di una separazione non lineare tra classi (Fig. 1d), il perceptrone non è in grado di eseguire correttamente la classificazione. Un caso emblematico è quello dell'apprendimento delle funzioni logiche: come riconosciuto da Marvin L. Minsky e Seymour A. Papert già nel 1969 [5], il perceptrone è in grado di eseguire operazioni logiche linearmente separabili, come AND e OR, ma non quelle non-linearmente separabili, come la funzione XOR. Per questo tipo di classificazioni, è necessario aumentare la complessità della rete, aggiungendo alla rete neurale uno o più strati intermedi, definiti nascosti (*hidden layers* in inglese), a formare il cosiddetto perceptrone multistrato, o *multiple layer perceptron* (MLP) in inglese (Fig. 1e).

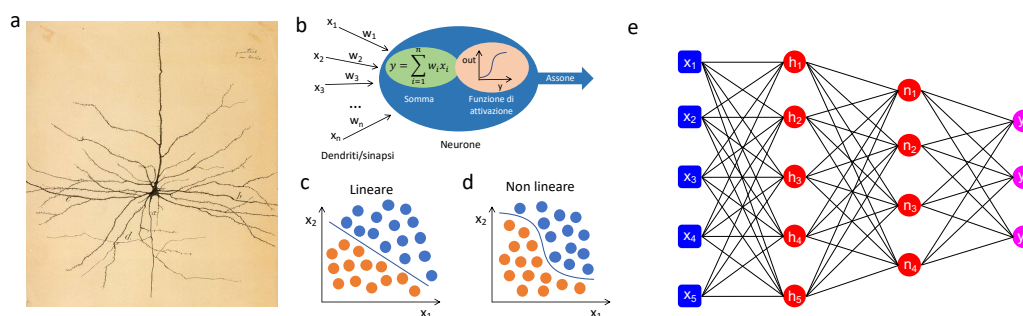


Figura 1

Alcune pietre miliari nella storia dell'intelligenza artificiale. (a) Schema di un neurone biologico di Cajal, (b) modello matematico di neurone secondo McCulloch e Pitts, riconoscimento con classi separabili (c) linearmente e (d) non linearmente, (e) schematico di un perceptrone a più strati (MLP)

La necessità di ricorrere al MLP per affrontare il riconoscimento di classi non separabili linearmente creò una generale disillusione nei confronti del concetto di rete neurale. La conseguenza di ciò fu un periodo di stagnazione nei programmi governativi e nei finanziamenti alla ricerca nell'ambito delle reti neurali in particolare negli USA, che durò per buona parte degli anni 1970 e 1980. Infatti, nonostante la rete MLP apparisse come uno strumento promettente per affrontare e generalizzare il problema del riconoscimento e della classificazione, non era ancora chiaro come sviluppare una metodologia di apprendimento automatico e indipendente dall'uomo.

L'interesse per le reti neurali si risvegliò a metà degli anni 1980, con l'affermarsi di una nuova tecnica di apprendimento supervisionato, nota con il nome di *backpropagation* (BP, traducibile con il termine 'propagazione all'indietro') [6,7]. La BP è una tecnica iterativa che permette di aggiornare i pesi sinaptici in modo da minimizzare la funzione 'costo', che generalmente descrive l'errore nel riconoscimento. Nella pratica, attraverso questa tecnica, si può addestrare un MLP presentando una sequenza di pattern, ad esempio immagini di cifre numeriche scritte a mano [8]. A ogni presentazione, viene anche valutato l'errore compiuto dai neuroni di output, dove, ad esempio, la presentazione della cifra '7' deve corrispondere ad un segnale massimo da parte del neurone che identifica la classe '7', e ad un segnale minimo da parte di tutti gli altri neuroni di classificazione. Ciascun errore viene poi fatto 'propagare all'indietro', cioè viene usato come coefficiente moltiplicativo per correggere ogni peso sinaptico della rete. Alla fine di questo apprendimento supervisionato con un numero adeguato di presentazioni di ogni classe, la rete riesce a raggiungere livelli di accuratezza nella classificazione molto alti, persino superiori a quelli umani [9].

Sebbene sia estremamente potente, la tecnica di BP è anche molto onerosa dal punto di vista computazionale, dato che l'apprendimento richiede numerose presentazioni, anche dell'ordine di centinaia di migliaia. Questo causò una nuova stagnazione delle reti neurali tra la fine degli anni 1990 e l'inizio degli anni 2010. Agli inizi degli anni 2010, tuttavia, il *deep learning* (termine che identifica l'uso di reti a multistrato per l'AI) conobbe una nuova fase di grande successo, dovuta alla disponibilità di macchine di calcolo in grado di elaborare una grande quantità di dati nell'unità di tempo. Questi calcolatori sono per lo più microprocessori grafici (*graphical processing unit*, GPU), inizialmente concepiti per l'elaborazione veloce delle immagini nei videogiochi, ma ben presto assurti a tecnologia di eccellenza per le applicazioni di AI. Un altro elemento abilitante per la diffusione del *deep learning* è stata anche la disponibilità di grandi database (Big Data) per l'addestramento, contenenti ad esempio immagini provenienti da Internet.

3. Tecnologie di deep learning

Oggigiorno, il *deep learning* costituisce l'ossatura di buona parte degli strumenti di AI, in gran parte implementati nei social network, nelle automobili a guida autonoma, e negli assistenti virtuali. Il riconoscimento di immagini, in particolare, è in gran parte affidato ad una specifica categoria di rete a multistrato, nota con il nome di rete convoluzionale (*convolutional neural network*, CNN) [10]. Nella

CNN, l'immagine in ingresso è scandita da una serie di filtri, ognuno ottimizzato allo scopo di rivelare la presenza di caratteri distintivi (*feature*), come linee, angoli, e altre forme più complesse. Il successo della CNN nel riconoscimento di immagini è da attribuire alla condivisione di molte sinapsi artificiali nella ricerca di *feature*, che permette di condurre sia l'apprendimento che il riconoscimento con un numero di sinapsi molto inferiori rispetto all'approccio MLP [7].

La CNN ha riscosso notevole successo nella classificazione di immagini, come i caratteri scritti a mano sugli assegni [10], e nel riconoscimento facciale. Lo strumento DeepFace di Facebook, ad esempio, che è basato su una rete a multistrato con 120 milioni di pesi, ha dimostrato un'accuratezza del 97.35% nel riconoscere i volti umani entro un insieme di 4000 individui [11]. La CNN può anche riconoscere oggetti in un generico contesto, come ad esempio un carro trainato da cavalli sulla strada (Fig. 2a), o un muletto all'esterno di un magazzino industriale (Fig. 2b). Questi risultati costituiscono i tasselli fondamentali per la cosiddetta visione artificiale o *computer vision*, che è una tecnologia abilitante per la robotica ed in particolare per la guida autonoma [12]. Un'automobile senza guidatore, ad esempio, deve automaticamente interpretare lo scenario che si presenta nella direzione di marcia per riconoscere la posizione della corsia, la segnaletica stradale, le altre automobili, i pedoni ed ogni possibile ostacolo (Fig. 2c).

Un altro strumento del *deep learning* di enorme interesse pratico è il riconoscimento vocale, che ad esempio riveste un ruolo chiave negli assistenti virtuali come Siri di Apple e Alexa di Amazon. La tecnologia *mainstream* per il riconoscimento vocale (*speech recognition*) è la *long short-term memory* (LSTM), che consiste in una rete ricorsiva in grado di memorizzare il segnale di ingresso e confrontarlo con i segnali agli istanti precedenti [13]. Mentre nelle reti come l'MLP o la CNN gli ingressi si propagano in una sola direzione, cioè dall'ingresso verso l'uscita dove avviene la classificazione, le reti ricorsive sono caratterizzate da un esteso feedback che permette di collocare la rete in un determinato 'stato'. Questo permette alle reti ricorsive di possedere una memoria che le rende in grado di effettuare applicazioni complesse, come il riconoscimento di un discorso, la traduzione in tempo reale, e l'elaborazione naturale del linguaggio.

Uno dei limiti fondamentali del *deep learning* è la necessità di ricorrere all'apprendimento supervisionato, che richiede estesi archivi di dati (database) istruiti da operatori umani per classificare ogni dato con la sua etichetta (*label*). Al contrario, l'apprendimento degli esseri umani e degli animali segue un approccio non supervisionato, dato che gran parte della nostra conoscenza si fonda più sull'esperienza diretta che sull'insegnamento di altri. Questo tipo di apprendimento trova espressione nella cosiddetta tecnica di apprendimento per rinforzo (*reinforcement learning*), che permette ad una rete a multistrato di imparare dall'esperienza, potenziando le sinapsi che contribuiscono ai propri successi e deprimendo quelle che invece portano agli insuccessi. Da questo punto di vista ha ricevuto notevole risonanza un recente esperimento di DeepMind, una start-up acquisita da Google nel 2014. DeepMind ha dimostrato che un calcolatore può imparare a giocare a 49 videogiochi Atari, raggiungendo

in più della metà dei giochi un'abilità superiore a quelle di giocatori professionisti [14]. In tutti i casi, la macchina è stata addestrata facendola giocare ripetutamente, in modo da apprendere abilità di gioco e 'trucchi' speciali sia dalle vittorie che dalle sconfitte, proprio come nella nostra esperienza di tutti i giorni. In particolare, si noti che la tecnica di apprendimento con rinforzo si ispira ad un sistema di premiazione riscontrato nel cervello umano, che si basa su un importante neurotrasmettitore chimico chiamato dopamina. La stessa tecnica è stata inoltre estesa ad obiettivi più complessi, quali l'antico gioco del Go [15]. Il programma AlphaGo, nome in codice del progetto di DeepMind per l'apprendimento di tale gioco, ha battuto nel 2016 l'allora campione mondiale Lee Sedol (Fig. 2d). Da notare che, al contrario di casi precedenti di successo, come la vittoria nel 1997 di Deep Blue di IBM nel gioco degli scacchi contro l'allora campione del mondo Garry Kasparov, l'apprendimento di AlphaGo era basato in gran parte sull'esperienza prova-ed-errore invece che su un programma rigidamente strutturato a priori.

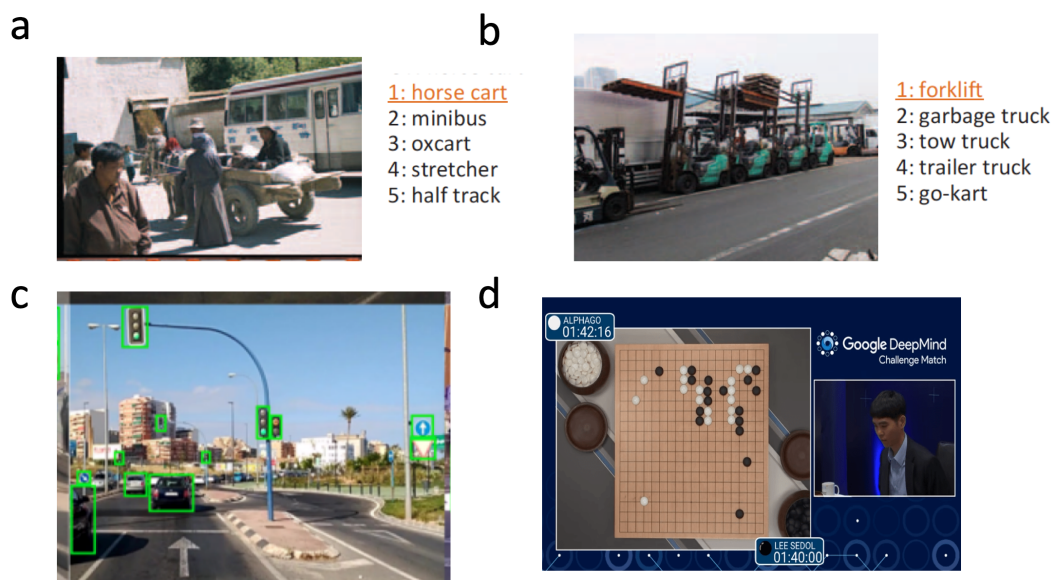


Figura 2

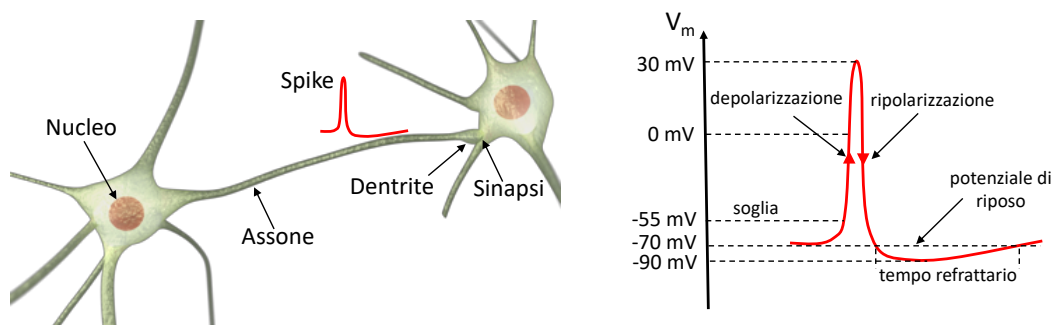
Stato dell'arte sul deep learning. Le reti neurali convoluzionali (CNN) adottate per il deep learning consentono di riconoscere oggetti in un generico contesto come, ad esempio, (a) un carro trainato da cavalli, o (b) un muletto industriale [9]. (c) Riconoscimento di oggetti in un frame visivo per la guida autonoma [12], e (d) gioco del Go mediante apprendimento con rinforzo (photo from Prachatai's photostream on flickr Creative Commons). Riprodotto da [9,12]. Copyright IEEE.

4. Oltre il deep learning: macchine ispirate alla mente umana

I successi del *deep learning* sono in gran parte attribuibili al connubio tra un potente algoritmo (l'apprendimento supervisionato), un'ampia disponibilità di dati per l'apprendimento, e grandi risorse di calcolo ad alte prestazioni, come la GPU. Sono in molti, tuttavia, a ritenere che quella del *deep learning* possa rivelarsi una 'bolla' che potrebbe presto rallentare il suo passo di sviluppo. Il principale limite è infatti la scarsa attinenza degli algoritmi di *deep learning* con la mente umana. Per esempio, un bambino di tre anni è in grado di riconoscere una tigre dopo averla vista la prima volta allo zoo, conservando tale esperienza per tutta la vita. Al contrario, le reti a multistrato supervisionate richiedono migliaia o addirittura milioni di presentazioni di immagini raffiguranti tigri con varie angolazioni, orientazione della luce, etc., per arrivare a sufficienti livelli di accuratezza. Se l'obiettivo è quindi quello di realizzare macchine autonome, in grado di interpretare il mondo che ci circonda e prendere decisioni immediate con alta efficienza energetica, potrebbe risultare più efficiente sviluppare algoritmi e architetture di calcolo che si ispirino direttamente al cervello umano. È questo l'obiettivo primario dei sistemi neuromorfici, cioè di quei sistemi cioè che si prefiggono di replicare il comportamento delle strutture neurobiologiche nel cervello umano.

Ancora oggi, la comprensione completa del funzionamento del cervello umano appare come una grande sfida. Alcuni tratti distintivi del cervello possono comunque essere presi a riferimento per costruire una macchina di calcolo ispirata ad esso. Prima di tutto, il cervello consta di una rete neurale, in cui non vi è alcuna separazione fisica tra unità di calcolo ed unità di memoria, come invece succede nei calcolatori digitali, ad esempio la GPU. Pertanto, la migliore strategia per imitare il cervello umano è quella di replicare la sua architettura già a partire dall'implementazione in hardware, ricreando una fitta rete di neuroni collegati tra loro da giunzioni sinaptiche [16]. Da questo punto di vista, una proprietà chiave del cervello umano è la sua alta *connettività*. A fronte di circa 10^{11} neuroni, il cervello umano conta infatti all'incirca 10^{15} sinapsi, che indica che in media, nel cervello, ogni neurone è collegato ad altri 10,000 neuroni [17]. Questa proprietà ha avuto un impatto decisivo sulla nostra evoluzione in quanto è proprio dall'altissima connettività neuronale del cervello umano che deriva la superiorità intellettuale della nostra specie sugli altri animali.

Sebbene da un certo punto di vista il cervello possa assomigliare alle reti neurali a multistrato, come ad esempio la rete MLP in Fig. 1e, la tipologia dei segnali trasmessi ed il metodo di apprendimento sono totalmente diversi. Nel cervello, infatti, i neuroni emettono impulsi elettrici (spike).

**Figura 3**

Disegno della struttura di un neurone biologico basata sull'assone e sulle dendriti, e del suo potenziale interno V_m che, superata una certa soglia, aumenta repentinamente causando l'emissione di un impulso elettrico o spike verso la sinapsi formata con la dendrite di un altro neurone.

Come mostrato in Fig. 3, il neurone mantiene normalmente un potenziale interno, o potenziale di membrana, V_m , costante attorno a -70 mV. Quando i neuroni contigui inviano stimolazioni attraverso le loro dendriti, il potenziale V_m del neurone aumenta. Se V_m supera una determinata soglia, attorno a -55 mV, si verifica un evento di *spike*, dove V_m aumenta repentinamente fino a circa 30 mV, per poi ridiscendere a circa -90 mV. Lo *spike* dura generalmente qualche millisecondo, ed è necessario un tempo refrattario di qualche millisecondo perchè V_m ritorni al valore stazionario iniziale di -70 mV. Si noti che la natura elettrica di questo spike ha origine dai processi biologici di depolarizzazione e ripolarizzazione, che sono caratterizzati rispettivamente dall'ingresso di ioni sodio e dal rilascio di ioni potassio. Dopo aver percorso l'assone, cioè il terminale di uscita del neurone, lo *spike* raggiunge le sinapsi di collegamento con le dendriti di altri neuroni, dove può essere comunicato ed elaborato.

La natura impulsiva dello *spike* è uno dei segreti che garantiscono al cervello l'alta efficienza energetica che lo contraddistingue. Infatti, ogni *spike* costa l'equivalente di 2.4×10^9 molecole di adenosin-trifosfato (ATP), pari a circa 0.1 nJ [18]. Per minimizzare il consumo energetico, l'utilizzo di *spike* nel cervello è limitato alla stretta necessità, ad esempio quando si verifica un evento sensoriale come la visione di un oggetto o la percezione di un suono. L'attività dei neuroni viene quindi centellinata sia temporalmente, per effetto della comunicazione a *spike*, sia spazialmente, per effetto della codifica 'sparsa' legata all'elevata specializzazione di ciascun neurone [19]. Si noti infine che l'attività di generazione di *spike* nel cervello è per definizione *asincrona*, cioè non soggetta ad una sincronizzazione globale (come il segnale di clock in un calcolatore digitale), ma legata unicamente al verificarsi di un evento sensoriale. Ad esempio, un particolare evento, come il riconoscimento di un volto noto, attiverà una certa regione del nostro cervello, altrimenti inattiva, per indurre una reazione, come il nostro saluto. L'attivazione di una parte del cervello stimolata dall'evento permette di tenere attivi solo un numero limitato di neuroni in ogni istante, in modo da minimizzare il consumo energetico.

Riguardo la metodologia di apprendimento nel cervello umano e animale, questo avviene in gran parte senza supervisione. La tecnica di BP, ancorché estremamente efficace nell'addestramento di reti neurali per il *deep learning*, non ha infatti nulla a che vedere con i processi di apprendimento biologico. Questo avviene invece mediante complessi meccanismi biochimici a livello della sinapsi, cioè il ponte di collegamento tra l'assone di un neurone e la dendrite di un altro neurone ricevente. A livello biologico, l'apprendimento avviene mediante il meccanismo della *plasticità sinaptica*, dove le sinapsi possono essere potenziate, depresse, o addirittura create ex-novo attraverso processi biochimici e fisiologici ancora non del tutto compresi.

Uno dei meccanismi di apprendimento, osservato sperimentalmente nell'ippocampo, è la plasticità sinaptica dipendente dal tempo (*spike timing dependent plasticity*, STDP), dove il ritardo tra gli *spike* di due neuroni connessi da una sinapsi determina il potenziamento o la depressione della sinapsi stessa [20]. Quando il neurone pre-sinaptico emette uno *spike* prima di quello post-sinaptico, entro una finestra temporale di qualche decina di millisecondi, la sinapsi viene potenziata, cioè il prossimo *spike* pre-sinaptico verrà trasmesso a quello post-sinaptico con maggiore efficacia. Se invece è il neurone post-sinaptico a precedere quello pre-sinaptico, la sinapsi va incontro ad una depressione. Meccanismi di plasticità più complessi sono anche stati proposti, come la plasticità dipendente dalla frequenza degli *spike* [21], o il fenomeno del tripletto di *spike*, dove è la concomitanza di 3 *spike* a modificare il peso sinaptico [22]. Tali meccanismi sono alla base di processi di apprendimento non supervisionato caratteristici delle reti neurali [23].

5. I circuiti neuromorfici

Un circuito neuromorfico si propone di riprodurre le fondamentali proprietà di calcolo del cervello umano, quali la sua architettura a rete neurale sparsa ed altamente connessa, la natura impulsata (*spiking*) e asincrona dell'informazione, e la plasticità sinaptica. Fin dagli anni 1990, l'ingegneria neuromorfica si è orientata verso circuiti analogici, per meglio riprodurre le proprietà asincrone e continue del segnale elettrico di *spike*. Il pioniere in questo campo è stato Carver Mead del California Institute of Technology (Caltech), esperto di progettazione analogica, e fautore di una tecnologia di circuiti neuromorfici basata sul utilizzo di transistori polarizzati in nella regione di funzionamento sottosoglia. In questa modalità, la tensione di gate che comanda la corrente del transistor è inferiore alla tensione di soglia, in modo da limitare la corrente tra 1 nA e 1 μ A. La polarizzazione in sottosoglia ha il pregio, quindi, di minimizzare il consumo di energia e al tempo stesso di poter meglio imitare i meccanismi di diffusione ionica presenti nel neurone e nella sinapsi, dato che gli elettroni lungo il canale di un transistor in sottosoglia si muovono anch'essi per diffusione. La scuola di Mead ha sviluppato i primi concetti di neuroni *spiking* e di sinapsi capaci di apprendere, il tutto mediante circuiti analogici integrati nel silicio. Mead è stato anche il primo scienziato a introdurre il concetto di circuito *neuromorfico*, per esprimere l'obiettivo di imitare la rete neurale biologica fin dalla sua architettura e dal suo *modus operandi* [24].

Con il progredire delle tecnologie di integrazione microelettronica, le implementazioni hardware di circuiti neuromorfici sono state declinate in diverse soluzioni, che possono oggi essere riassunte in 4 diversi approcci progettuali. Da una parte, il circuito TrueNorth realizzato da IBM (Fig. 4a) consiste in un circuito *multicore* totalmente digitale, con un milione di neuroni e 256 milioni di sinapsi integrate in un singolo chip [25]. Il nocciolo (*core*) nell'architettura di TrueNorth presenta 256 ingressi (assoni) e 256 uscite (neuroni) collegati da 256x256 connessioni sinaptiche completamente riconfigurabili, in un'architettura dove per la prima volta la memoria e l'elaborazione dei dati sono colocalizzati nella stessa area del silicio. Questo rappresenta la principale novità rispetto all'architettura tradizionale di von Neumann, dove invece la memoria e l'unità di elaborazione centrale (*central processing unit*, CPU) sono realizzati su chip fisicamente distinti. La chiave per l'alta efficienza di elaborazione diventa quindi la massiccia interconnessione tra i vari core neurosinaptici, che avviene mediante l'algoritmo della rappresentazione indirizzo-evento (*address-event representation*, AER) [26]. Nell'AER, le informazioni tra un core e l'altro vengono trasmesse mediante pacchetti di bit contenenti l'indirizzo del neurone che ha emesso uno *spike*. In questo modo, è possibile contenere al minimo la complessità delle interconnessioni tra i vari core neuromorfici, pur mantenendo un'alta densità di *spike* trasmessi. Il circuito digitale è sincrono, con un clock globale di frequenza 1 kHz, ed è stato integrato nel silicio con la tecnologia 28 nm di Samsung.

Un altro esempio di implementazione puramente digitale è il circuito SpiNNaker dell'Università di Manchester (Fig. 4b) [27]. Spinnaker è stato sviluppato assemblando 18 microprocessori ARM, all'interno dei quali vengono simulati gli *spike* neuronali trasmessi all'interno della rete. La trasmissione avviene in modo asincrono per una migliore attinenza con quanto avviene nella corteccia cerebrale. L'approccio modulare permette di riconfigurare sia il numero di neuroni simulati all'interno del singolo core, sia il numero di core, che ha raggiunto il milione nell'implementazione più recente [28].

Il circuito BrainScales dell'Università di Heidelberg propone un'implementazione mista, dove il neurone è descritto da un circuito analogico mentre la comunicazione tra i neuroni è affidata ad un approccio digitale. Il circuito BrainScales permette l'assemblaggio modulare su scala di wafer, ad esempio collegando 20 wafer da 8 pollici per raggiungere un totale di 4 milioni di neuroni (Fig. 4c) [29,30]. La computazione neurale in BrainScales può essere accelerata di 10,000 volte, in modo da velocizzare enormemente la simulazione di un processo cognitivo ai fini della comprensione dei meccanismi di elaborazione nel cervello. La simulazione di reti neuro-biologiche è infatti uno degli obiettivi cardine del progetto europeo Human Brain, all'interno del quale sono stati sviluppati i circuiti SpiNNaker e BrainScales.

Uno dei limiti dei precedenti circuiti è la difficoltà nel descrivere l'apprendimento, che è l'aspetto caratterizzante del cervello umano. Il processore neuromorfico ROLLS dell'Università di Zurigo (Fig. 4d) riunisce alcuni punti salienti dei circuiti precedenti, come l'elaborazione mista analogico-digitale, il protocollo di comunicazione AER e l'architettura riconfigurabile non-von Neumann, offrendo in più, per la prima volta, la plasticità sinaptica secondo

algoritmi ispirati all'STDP. La plasticità sinaptica abilita ROLLS all'emulazione di meccanismi di apprendimento tipici del cervello umano, come la memoria associativa nelle reti ricorsive di Hopfield [16], che è alla base dei processi di orientamento per la navigazione di agenti autonomi, come robot e droni [31].

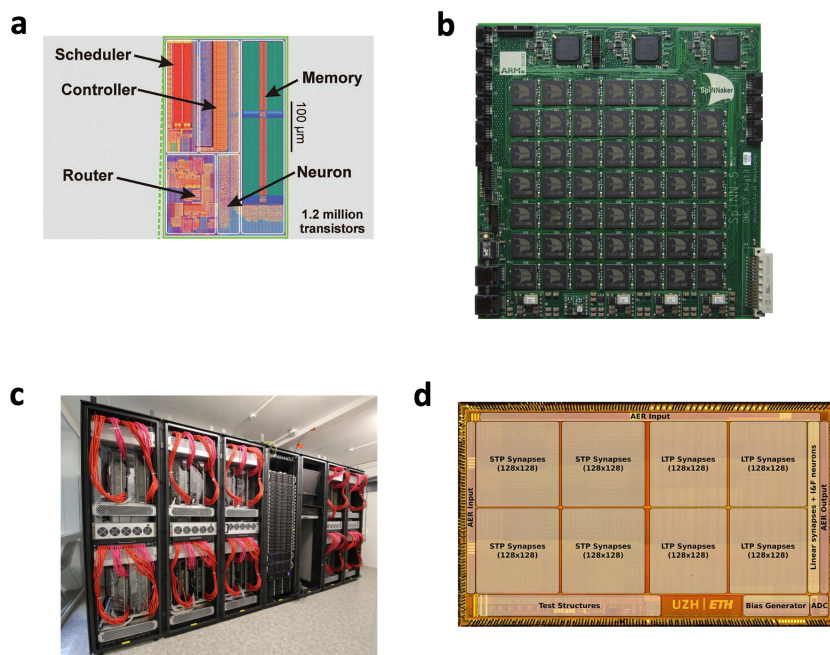


Figura 4

Recenti implementazioni circuitali in tecnologia CMOS di reti neurali complesse. (a) Schematico di un core neurosinaptico alla base del chip TrueNorth realizzato da IBM [25], (b) scheda che assembla fino a 48 circuiti SpiNNaker [27], (c) sistema BrainScales basato su 20 wafer [29] e (d) foto del processore neuromorfico ROLLS dell'Università di Zurigo [16].

Sebbene la funzionalità dei processori neuromorfici sia molto più vicina al comportamento delle reti neurali biologiche, le applicazioni per questo tipo di circuiti non sono ancora chiare. Da un lato, sono in molti a sperare di poter utilizzare i circuiti neuromorfici per meglio apprendere i meccanismi di calcolo e di apprendimento che caratterizzano il cervello umano. È questa infatti una delle più forti motivazioni alla base di progetti pluriennali quali lo Human Brain Project [32]. Grazie alla somiglianza con il cervello umano, i circuiti neuromorfici potrebbero anche essere impiegati per una elaborazione di informazioni più veloce e con una migliore efficienza energetica. In questo ambito, le migliori prestazioni si ottengono quando il sistema neuromorfico si interfaccia direttamente con sensori spiking, come i sistemi a visione differenziali (dynamic vision sensor, DVS) [33]. In un DVS, invece di registrare l'intensità della luce pixel per pixel come nelle telecamere convenzionali, vengono rivelati solo gli eventi temporali, come le variazioni di intensità. In questo modo è possibile ridurre notevolmente il volume di informazioni trasmesse, che nel DVS si limitano

alle coordinate dell'evento all'interno del frame, invece che il frame completo dell'immagine. È stato dimostrato che i circuiti neuromorfici, combinati con i sistemi event-driven come il DVS, permettono di accelerare enormemente l'elaborazione delle informazioni, consentendo di riconoscere e seguire un oggetto veloce in tempo reale [34].

Infine, le reti neuromorfiche sono altamente promettenti nella sfera del computing, in particolare nell'ambito del calcolo stocastico [35]. I circuiti neuromorfici sono infatti in grado di risolvere problemi particolarmente ardui per un computer convenzionale, come la risoluzione di un Sudoku [36] o di altri problemi di soddisfazione di vincoli (constraint satisfaction problem, CSP) [37]. A sottolineare la novità dell'approccio neuromorfico e la sua importanza per il computing a bassa energia, anche alcuni player del calibro di Intel hanno recentemente annunciato lo sviluppo di reti neurali spiking per applicazioni generiche [38].

6. I nuovi dispositivi neuromorfici

Uno dei limiti attuali dei circuiti neuromorfici riguarda la tecnologia convenzionale dei transistori CMOS (complementary metal-oxide-semiconductor), che potrebbe avere difficoltà a fornire la necessaria densità di sinapsi e a riprodurre la funzionalità delle reti neurobiologiche con circuiti sufficientemente semplici. La realizzazione di una sinapsi in tecnologia CMOS, infatti, richiede una dozzina di transistori e un paio di condensatori, per realizzare le funzioni temporali tipiche della plasticità a breve termine [39]. Un numero doppio di transistori è invece richiesto per un neurone analogico di tipo integrate-and-fire [39]. Ciò è dovuto alla relativa semplicità del transistor rispetto alla complessità delle funzioni neuromorfiche di un neurone o di una sinapsi in biologia. Questo problema è all'origine del limitato numero di neuroni e sinapsi che si possono realizzare su un chip di area medio piccola, ad esempio circa 256 neuroni e 133 mila sinapsi su un chip di 51.4 mm² in tecnologia 180 nm [16]. Per raggiungere un numero più elevato di neuroni è necessario ricorrere all'integrazione sulla scala del wafer, che tuttavia è incompatibile con tutte le applicazioni dove si desidera un sistema neuromorfico di piccolo ingombro e basso consumo [40].

Per superare le limitazioni imposte dalla tecnologia CMOS, è in atto un'intensa ricerca sull'integrazione di nuove tipologie di dispositivi nanoelettronici in circuiti neuromorfici. La tecnologia più promettente è quella delle cosiddette memorie emergenti, anche note con il nome di memristori. Dispositivi come le memorie a switching resistivo (resistive-switching random access memory, RRAM), le memorie a cambiamento di fase (phase change memory, PCM), le memorie magnetoresistive (magnetic random access memory, MRAM) e le memorie ferroelettriche (ferroelectric random access memory, FERAM) offrono una capacità di miniaturizzazione superiore al transistor CMOS, pertanto sono estremamente attraenti per tutte le applicazioni di computing in memoria (Fig. 5) [41]. Ad esempio, i dispositivi RRAM e PCM possono raggiungere dimensioni attorno ai 10 nm, che non sono accessibili alle memorie convenzionali CMOS [42]. L'estrema miniaturizzazione, unita alla possibilità di realizzare operazioni algebriche complesse come il prodotto matrice-vettore direttamente nella

matrice di memoria [41], rende le nuove memorie particolarmente attraenti per le applicazioni di deep learning [43].

Aldilà delle dimensioni estremamente scalabili, le memorie emergenti presentano anche un'interessante molteplicità di fenomeni fisici che possono essere sfruttati per realizzare funzioni neuromorfiche complesse. Innanzitutto, è stato dimostrato che i dispositivi RRAM permettono di imitare la plasticità sinaptica di tipo STDP, dove il ritardo tra gli spike pre-sinaptici e post-sinaptici stabilisce la variazione del peso sinaptico [44]. La Fig. 6 mostra il processo di addestramento non supervisionato di una rete neurale con 16 sinapsi RRAM, dove la presentazione di 3 pattern in sequenza (Fig. 6a-c) porta all'apprendimento automatico degli stessi grazie all'STDP (Fig. 6d-f) [44]. In alcuni tipi di dispositivi RRAM, noti come memristori di secondo ordine [45], si è evidenziata la possibilità di una memoria a breve termine, in cui il dispositivo può ricordare per un breve tempo, dell'ordine di pochi microsecondi, l'applicazione di un precedente impulso (Fig. 7a). [21]. Come mostrato in Fig. 7b, l'applicazione di 2 impulsi sufficientemente separati nel tempo non modifica il dispositivo, mentre l'applicazione di due impulsi in rapida successione comporterà l'aumento di conduttanza, come in una sinapsi biologica. Questo permette di replicare a livello hardware la plasticità di tipo STDP sfruttando direttamente le proprietà fisiche peculiari del dispositivo RRAM. Un simile comportamento può essere utile anche per emulare neuroni integrate-and-fire, con il vantaggio di poter rimpicciolire notevolmente l'ingombrante circuito neuronale grazie alle proprietà fisiche peculiari del dispositivo [41].

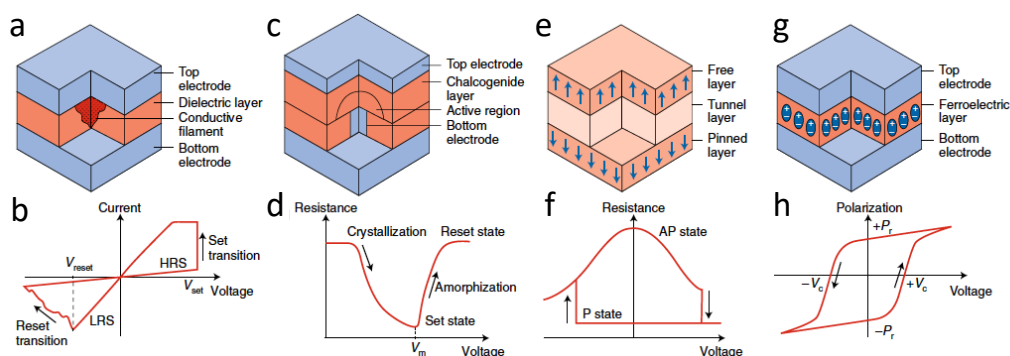
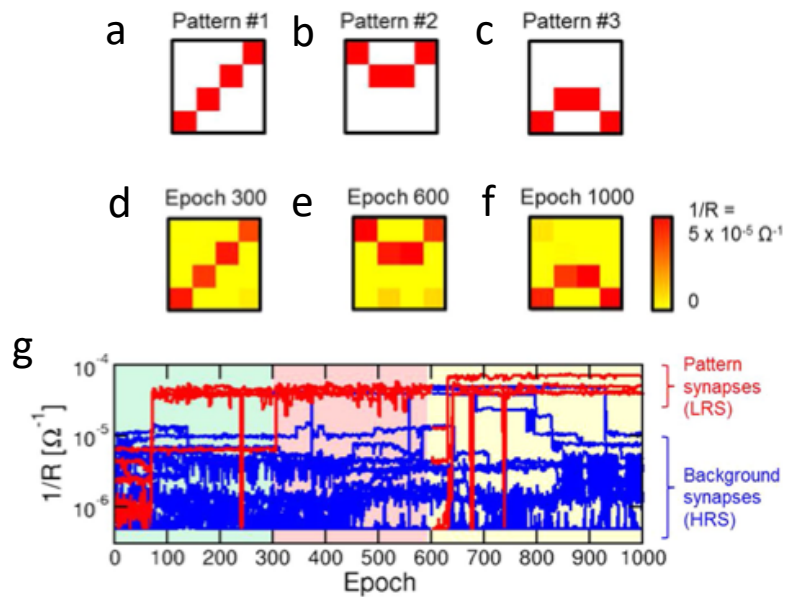
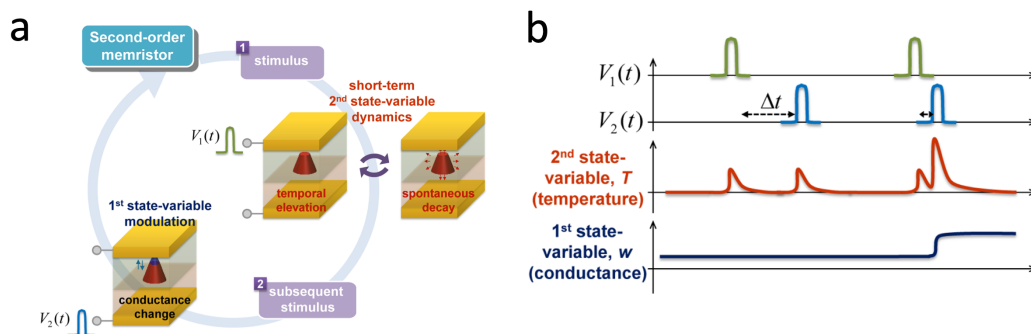


Figura 5

Principali tipologie di memorie non-volatili emergenti racchiuse sotto il nome di memristori. (a) Struttura di una memoria a switching resistivo (RRAM) con (b) la rispettiva caratteristica di funzionamento corrente-tensione. (c) Struttura di una memoria a cambiamento di fase (PCM) con (d) la rispettiva caratteristica che mostra l'andamento della resistenza in funzione della tensione applicata ai capi del dispositivo (R-V). (e) Struttura di una memoria magnetoresistiva (MRAM) con (f) la corrispondente caratteristica R-V. (g) Struttura di una memoria ferroelettrica con (h) la rispettiva caratteristica, che mostra la polarizzazione del dispositivo in funzione della tensione applicata [41]. Copyright Springer Nature Publishing AG.

**Figura 6**

Descrizione dell'apprendimento non supervisionato basato sull'STDP in una rete neurale spiking con 16 sinapsi memristive RRAM. (a-c) Rappresentazione delle immagini con 4x4 pixel fornite in sequenza allo strato di neuroni di ingresso della rete. (d-f) Rappresentazione della conduttanza elettrica dei pesi sinaptici della rete misurata alla fine di ciascuna fase dell'esperimento, che evidenzia l'apprendimento di ciascuna immagine. (g) Evoluzione temporale della conduttanza dei pesi sinaptici, che mostra il potenziamento delle sinapsi stimolate dall'immagini fornite in ingresso e la depressione delle sinapsi non stimolate in ciascuna fase dell'esperimento. Ciò conferma la capacità di apprendimento non supervisionato di una rete neurale spiking dotata di sinapsi memristive RRAM. Adattata da [44].

**Figura 7**

Descrizione del processo di STDP in memristori di secondo ordine. (a) Schema della memoria a breve termine nel dispositivo, dove l'applicazione di un impulso viene 'ricordato' per un breve tempo, dell'ordine di qualche microsecondo, dal dispositivo, durante il quale il dispositivo può essere sensibile all'applicazione di un secondo impulso. (b) Schema dell'STDP, dove l'applicazione di 2 impulsi con grande separazione temporale non modifica la sinapsi, mentre una breve separazione temporale comporta il potenziamento sinaptico. Adattata da [45].

Un altro aspetto peculiare di questi dispositivi di memoria è il loro comportamento stocastico, che ancora una volta riflette alcuni tipici comportamenti neuro-biologici. Fig. 8a mostra il tipico segnale spiking di un neurone, dove il tempo tra 2 spike è normalmente casuale. Alcuni tipi di memoria permettono un simile andamento casuale di spike, come ad esempio le MRAM superparamagnetiche [46]. In questi dispositivi, simili a quelli di Fig. 5e, la riduzione delle dimensioni fa diminuire la barriera energetica tra i due stati di memoria, i cosiddetti stati parallelo e anti-parallelo. Il risultato è che il dispositivo oscilla spontaneamente a temperatura ambiente tra i due stati di memoria, in modo stocastico e con frequenza media controllabile dalla corrente di polarizzazione (Fig. 8b e c). Questo comportamento si è rivelato utile per codificare ed elaborare l'informazione mediante una popolazione di neuroni stocastici, in analogia con quanto avviene ad esempio nella corteccia visiva [47]. Il carattere stocastico appare quindi una delle prerogative uniche dei dispositivi di memoria che permettono di trasformare una caratteristica normalmente indesiderata, come la variabilità statistica, in una proprietà abilitante per l'elaborazione cognitiva. Lo studio di nuovi materiali, nuovi dispositivi, e dei relativi meccanismi fisici, risulta pertanto una carta vincente per lo sviluppo di circuiti neuromorfici che imitino il cervello umano sia nei processi elementari, sia nell'elevata densità di neuroni e sinapsi.

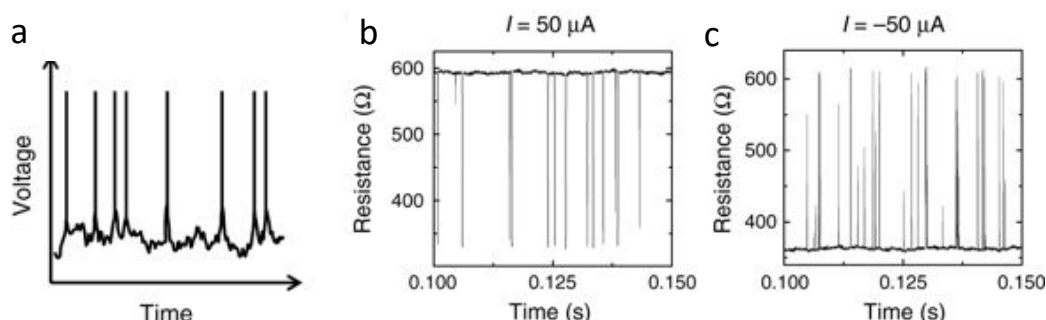


Figura 8

Descrizione della natura intimamente stocastica dell'attività di un neurone biologico evidenziata (a) dal caratteristico segnale impulsato di tensione generato in funzione del tempo. Tale comportamento può essere riprodotto utilizzando particolari dispositivi di memoria come le memorie superparamagnetiche, le quali oscillano in modo spontaneo tra due stati resistivi distinti (b,c) con una frequenza regolabile mediante la corrente di polarizzazione applicata. Adattata da [46].

7. Conclusione

I sistemi di calcolo neuromorfo, pensati come unità autonome in grado di riprodurre i meccanismi cognitivi del cervello umano, potrebbero rivoluzionare la nostra esistenza in molti aspetti. L'idea di disporre di assistenti personali veramente autonomi potrebbe facilitare la nostra vita nel lavoro e nel tempo libero. Sistemi in grado di monitorare ambienti e riconoscere volti e situazioni potrebbero migliorare la sicurezza dei luoghi pubblici, delle case, e delle

persone. La capacità di imparare dall'esperienza potrebbe permettere ai sistemi neuromorfici di migliorare giorno dopo giorno, superando addirittura i modelli umani in settori cruciali come l'ingegneria, la finanza, la politica e la medicina.

Per ora tutte queste potenziali innovazioni rimangono una visione attraente ma ancora utopistica. La loro realizzazione richiede ancora passi avanti significativi da parte dell'ingegneria neuromorfica e delle neuroscienze. Esiste infatti un'enorme lacuna nella nostra comprensione del cervello umano, che può essere colmata soltanto da anni di ricerca teorica e sperimentale. Gli stessi circuiti neuromorfici potrebbero essere di aiuto in questa ricerca: infatti, la diversa risposta di un sistema neuromorfico a seconda dei parametri impostati nel circuito (connettività neuronale, forma degli spike, plasticità) potrebbe permetterci di comprendere il modo in cui il cervello funziona, come in un esperimento di ingegneria inversa (reverse engineering). Un approccio congiunto dove si studi sia l'esperimento simulato in hardware, sia il suo corrispettivo biologico mediante esperimenti in vivo rappresenta un potente strumento per progredire nella conoscenza del cervello. Da questo punto di vista, un possibile percorso di ricerca è quello di prendere a riferimento cervelli meno evoluti di quello umano, che potrebbero anche essere più facilmente riprodotti nell'hardware. Ad esempio, il cervello di un'ape contiene 'solo' un milione di neuroni, che è una scala facilmente accessibile agli odierni prototipi neuromorfici.

Gli sviluppi futuri non verranno solo da una migliore comprensione del cervello, ma anche da tecnologie microelettroniche più evolute in termini di nuovi dispositivi, nuove architetture e nuovi processi di integrazione. Da questo punto di vista, la ricerca di materiali e dispositivi con proprietà 'neuromorfiche' è uno degli ambiti di ricerca più affascinanti. Dato che l'apprendimento è la proprietà più caratterizzante del cervello, sono i dispositivi di memoria ad essere oggetto della ricerca più intensa. Queste nuove tecnologie sono estremamente promettenti nell'offrire sia una notevole capacità di scaling, sia un 'portafoglio' di proprietà interessanti e uniche per abilitare processi neurali nel circuito. Infine, la possibilità di integrare questi dispositivi in tre dimensioni, come nella tecnologia 3D crosspoint recentemente introdotta nel mercato [48], potrebbe accelerare il passo verso reti neurali fisicamente connesse come nel cervello biologico.

Nel complesso, il mondo dell'ingegneria neuromorfica appare come uno dei più avvincenti nel panorama della ricerca ingegneristica, perché in esso convivono connotazioni di scienza 'hard' (nuovi materiali, tecnologia microelettronica ad alta densità, ingegneria di sistemi di elaborazione) e implicazioni biologiche, umanistiche, e persino filosofiche ed etiche. Un tale scenario crea un panorama complesso e articolato che presenta sfide da diversi punti di vista, che la comunità scientifica ha il compito di svelare in tutte le sue sfaccettature a vantaggio della società.

Ringraziamenti

L'autore ringrazia Valerio Milo e Wei Wang per la rilettura critica del manoscritto e il supporto grafico. Questo articolo ha ricevuto il finanziamento dell'European Research Council (ERC) nell'ambito del Programma di Ricerca e Innovazione Horizon 2020 dell'Unione Europea (grant 648635).

Bibliografia

- [1] Waldrop, M. M. (2013). "Smart connections", *Nature*, 503, 22-24.
- [2] McCulloch, W. S., Pitts, W. A. (1943). "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, 5, 115-133.
- [3] Hodgkin A. L., Huxley A. F. (1952). "The components of membrane conductance in the giant axon of Loligo", *The Journal of Physiology*, 116, 473-496.
- [4] Rosenblatt, F. (1957). *The Perceptron: A perceiving and recognizing automaton*, Report 85-460-1, Cornell Aeronautical Laboratory, Buffalo, New York.
- [5] Minsky, M. L., Papert, S. A. (1972). *Perceptrons: An introduction to computational geometry*, The MIT Press, Cambridge MA.
- [6] LeCun, Y. (1985). "Une procédure d'apprentissage pour réseau a seuil asymmetrique (A Learning Scheme for Asymmetric Threshold Networks)", *Proceedings of Cognitiva*, 85, 599-604.
- [7] LeCun, Y., Bengio, Y., Hinton, G. (2015). "Deep learning", *Nature*, 521, 436-444.
- [8] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). "Gradient-based learning applied to document recognition", *Proceedings of IEEE*, 86, 2278-2324.
- [9] He, K., Zhang, X., Ren, S., Sun, J. (2015). "Delving Deep into Rectifiers: Surpassing Human-level Performance on ImageNet Classification", *IEEE International Conference on Computer Vision (ICCV)*, 1026-1034.
- [10] LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., Jackel, L. D. (1990). "Handwritten digit recognition with a back-propagation network" in *Advances in Neural Information Processing Systems (NIPS 1989)*, Denver (CO) (Vol. 2), Morgan Kaufmann, 396-404.
- [11] Taigman, Y., Yang, M., Ranzato, M., Wolf, L. (2014). "DeepFace: Closing the gap to human-level performance in face verification", *IEEE Conference on Computer Vision and Pattern Recognition*, 1701-1708.
- [12] Dominguez-Sanchez, A., Cazorla, M., Orts-Escolano, S. (2018). "A New Dataset and Performance Evaluation of a Region-Based CNN for Urban Object Detection" *Electronics* 7, 301.
- [13] Hochreiter, S., Schmidhuber, J. (1997). "Long short-term memory", *Neural Computation*, 9, 1735-1780.
- [14] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D. (2015). "Human-level control through deep reinforcement learning", *Nature*, 518, 529-533.
- [15] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T.,

- Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D. (2016). "Mastering the game of Go with deep neural networks and tree search", *Nature*, 529, 484-489.
- [16] Indiveri, G., Liu, S.-C. (2015). "Memory and information processing in neuromorphic systems", *Proceedings of the IEEE*, 103, 1379-1397.
- [17] Ho, V. M., Lee, J.-A., Martin, K. C. (2011). "The cell biology of synaptic plasticity", *Science*, 334, 623-628.
- [18] Lennie, P. (2003). "The cost of cortical computation", *Current Biology*, 13, 493-497.
- [19] Olshausen, B. A., Field, D. J. (2004). "Sparse coding of sensory inputs", *Current Opinion in Neurobiology*, 14, 481-487.
- [20] Bi, G.-Q., Poo, M.-M. (1998). "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type", *The Journal of Neuroscience*, 18, 10464-10472.
- [21] Bienenstock, E. L., Cooper, L. N., Munro, P. W. (1982). "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex", *The Journal of Neuroscience*, 2, 32-48.
- [22] Pfister, J. P., Gerstner, W. (2006). "Triplet of spikes in a model of spike timing-dependent plasticity", *The Journal of Neuroscience*, 26, 9673-9682.
- [23] Masquelier, T., Thorpe, S. J. (2007). "Unsupervised Learning of Visual Features through Spike Timing Dependent Plasticity", *PLoS Comput Biol*, 3(2): e31.
- [24] Mead, C. (1989). *Analog VLSI and Neural Systems*, Addison-Wesley, Boston, MA.
- [25] Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., Jackson, B. L., Imam, N., Guo, C., Nakamura, Y., Brezzo, B., Vo, I., Esser, S. K., Appuswamy, R., Taba, B., Amir, A., Flickner, M. D., Risk, W. P., Manohar, R., Modha, D. S. (2014). "A million spiking-neuron integrated circuit with a scalable communication network and interface", *Science*, 345, 668-673.
- [26] Mahowald, M. (1994). *An Analog VLSI System for Stereoscopic Vision*, Kluwer Academic, Boston.
- [27] Furber, S. B., Galluppi, F., Temple, S., Plana, L. A. (2014). "The SpiNNaker project", *Proceedings of the IEEE*, 102, 652-665.
- [28] <http://www.eenewsanalog.com/news/spinnaker-neuromorphic-supercomputer-reaches-one-million-cores> (ultimo accesso 29 Novembre 2018).
- [29] Furber, S. (2016). "Large-scale neuromorphic computing systems", *Journal of Neural Engineering*, 13, 051001
- [30] <https://spectrum.ieee.org/computing/hardware/the-brain-as-computer-bad-at-math-good-at-everything-else> (ultimo accesso 29 Novembre 2018).
- [31] Kreiser, R., Aathmani, D., Qiao, N., Indiveri, G., Sandamirskaya, Y. (2018). "Organizing sequential memory in a neuromorphic device using dynamic neural fields", *Frontiers in Neuroscience*, 12, 717.
- [32] <https://www.humanbrainproject.eu/en/> (ultimo accesso 29 Novembre 2018).

- [33] Lichtsteiner, P., Posch, C., Delbruck, T. (2008). "A 128×128 120 dB 15 μ s latency asynchronous temporal contrast vision sensor", *IEEE Journal of Solid-State Circuits*, 43, 566-576.
- [34] Serrano-Gotarredona, R., Oster, M., Lichtsteiner, P., Linares-Barranco, A., Paz-Vicente, R., Gómez-Rodríguez, F., Camuñas-Meza, L., Berner, R., Rivas-Perez, M., Delbruck, T., Liu, S.-C., Douglas, R., Häfliger, P., Jiménez-Moreno, G., Civit Balcells, A., Serrano-Gotarredona, T., Acosta-Jiménez, A. J., Linares-Barranco, B. (2009). "CAVIAR: A 45k neuron, 5M synapse, 12G connects/s AER hardware sensory-processing-learning-actuating system for high-speed visual object recognition and tracking", *IEEE Transactions on Neural Networks*, 20, 1417-1438.
- [35] Maass, W. (2014). "Noise as a resource for computation and learning in networks of spiking neurons", *Proceedings of the IEEE*, 102, 860-880.
- [36] Hopfield, J. J. (2008). "Searching for Memories, Sudoku, Implicit Check Bits, and the Iterative Use of Not-Always-Correct Rapid Neural Computation", *Neural Computation*, 20, 1119-1164.
- [37] Mostafa, H., Müller, L. K., Indiveri, G. (2015). "An event-based architecture for solving constraint satisfaction problems", *Nat. Commun.* 6, 8941.
- [38] Davis, M., Srinivasa, N., Lin, T.-H., Chinya, G., Cao, Y., Choday, S. H., Dimou, G., Joshi, P., Imam, N., Jain, S., Liao, Y., Lin, C.-K., Lines, A., Liu, R., Mathaikutty, D., McCoy, S., Paul, A., Tse, J., Venkataramanan, G., Weng, Y.-H., Wild, A., Yang, Y., Wang, H. (2018). "Loihi: A neuromorphic manycore processor with on-chip learning", *IEEE Micro*, 38, 82-99.
- [39] Chicca, E., Stefanini, F., Bartolozzi, C., Indiveri, G. (2014). "Neuromorphic electronic circuits for building autonomous cognitive systems", *Proceedings of the IEEE*, 102, 1367-1388.
- [40] Schemmel, J., Brüderle, D., Gribbl, A., Hock, M., Meier, K., Millner, S. (2010). "A wafer-scale neuromorphic hardware system for large-scale neural modeling", *Proceedings of International Symposium on Circuits and Systems*, 1947-1950.
- [41] Ielmini, D., Wong, H.-S. P. (2018). "In-memory computing with resistive switching devices", *Nature Electronics* 1, 333-343.
- [42] Govoreanu, B., Kar, G. S., Chen, Y.-Y., Paraschiv, V., Kubicek, S., Fantini, A., Radu, I. P., Goux, L., Clima, S., Degraeve, R., Jossart, N., Richard, O., Vandeweyer, T., Seo, K., Hendrickx, P., Pourtois, G., Bender, H., Altimime, L., Wouters, D. J., Kittl, J. A., Jurczak, M. (2011). "10×10 nm² Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation", *IEEE International Electron Devices Meeting (IEDM)*, 729-732.
- [43] Ambrogio, S., Narayanan, P., Tsai, H., Shelby, R. M., Boybat, I., di Nolfo, C., Sidler, S., Giordano, M., Bodini, M., Farinha, N. C. P., Killeen, B., Cheng, C., Jaoudi, Y., Burr, G. W. (2018). "Equivalent-accuracy accelerated neural network training using analogue memory", *Nature*, 558, 60-67.
- [44] Pedretti, G., Milo, V., Ambrogio, S., Carboni, R., Bianchi, S., Calderoni, A., Ramaswamy, N., Spinelli, A. S., Ielmini, D. (2017). "Memristive neural network for

on-line learning and tracking with brain-inspired spike timing dependent plasticity," *Scientific Reports*, 7, 5288.

[45] Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski J. K., Aono, M. (2011). "Short-term plasticity and long-term potentiation mimicked in single inorganic synapses", *Nature Materials*, 10, 591-595.

[46] Mizrahi, A., Hirtzlin, T., Fukushima, A., Kubota, H., Yuasa, S., Grollier, J., Querlioz, D. (2018). "Neural-like computing with populations of superparamagnetic basis functions", *Nature Communications*, 9, 1533.

[47] Pasupathy, A., Connor, C. E. (2002). "Population coding of shape in area V4", *Nature Neuroscience*, 5, 1332-1338.

[48] <https://www.intel.it/content/www/it/it/architecture-and-technology/intel-micron-3d-xpoint-webcast.html> (ultimo accesso 29 Novembre 2018).

Tabella degli acronimi

Acronimo	Definizione	Significato
AI	Artificial intelligence	Intelligenza artificiale
MLP	Multiple layer perceptron	Percettone multistrato, cioè una rete neurale che comprende uno o più strati, definiti nascosti (hidden layers in inglese), intermedi tra i terminali di ingresso e i neuroni di uscita.
BP	Backpropagation	Propagazione all'indietro, cioè quella tecnica iterativa che permette di aggiornare i pesi sinaptici di una rete neurale in modo da minimizzare la funzione 'costo', che generalmente coincide con l'errore a livello dei neuroni di uscita, o di classificazione.
GPU	Graphical processing unit	Unità di elaborazione grafica, cioè un microprocessore per l'elaborazione veloce di immagini, specialmente applicato nei videogiochi e più recentemente nel training di reti neurali di tipo MLP.
CNN	Convolutional neural network	Rete neurale convoluzionale, cioè una rete in cui vi è uno o più strati dove il segnale viene posto in convoluzione con un filtro fisso, al fine di estrarre delle caratteristiche (feature) dal pattern di ingresso, come ad esempio linee, angoli e altre forme geometriche.
LSTM	Long short-term memory	Memoria a breve termine lunga, cioè una speciale rete ricorsiva che mantiene i dati e/o le feature per un certo lasso di tempo (memoria a breve termine) al fine di permettere la classificazione di pattern temporali, quali ad esempio sequenze audio.
ATP	Adenosin-trifosfato	Molecola responsabile della conversione di energia potenziale chimica nella quasi totalità delle reazioni metaboliche endoergoniche.
STDP	Spike timing dependent plasticity	Plasticità dipendente dalla tempistica degli spike, cioè un meccanismo di apprendimento che potenzia la sinapsi nel caso in cui lo spike presinaptico precede lo spike post-sinaptico, e deprime la sinapsi nel caso in cui lo spike presinaptico segue lo spike post-sinaptico.

CPU	Central processing unit	Unità di elaborazione centrale, cioè il microprocessore responsabile del calcolo digitale in tutti i computer e smartphone.
AER	Address-event representation	Rappresentazione per indirizzo-evento, cioè un protocollo di comunicazione all'interno di reti neurali spiking, dove ogni segnale di spike tra i vari core neuronali contiene l'indirizzo del neurone di partenza.
DVS	Dynamic vision sensor	Sensore di visione dinamico, detto anche retina artificiale, cioè un sensore che rivela ad ogni pixel le differenze tra l'intensità di luce ad un certo istante e l'intensità di luce ad un istante precedente. In questo modo, il sensore rivela solo eventi, cioè variazioni di intensità, in modo da limitare il numero di informazioni da comunicare al sistema neuromorfico di elaborazione.
CSP	Constraint satisfaction problem	Problema di soddisfacimento di vincoli, cioè una classe di problemi in cui si vogliono determinare tutte le possibili combinazioni di variabili di un certo sistema che soddisfino un certo set di vincoli, come ad esempio il problema delle otto regine ed il Sudoku.
CMOS	Complementary metal oxide semiconductor	Metallo-ossido-semiconduttore complementare, cioè la famiglia di porte logiche più diffusa nei circuiti integrati digitali.
RRAM	Resistive switching random access memory	Memoria a switching resistivo, cioè una memoria a due terminali la cui resistenza può essere modificata dall'applicazione di impulsi esterni mediante la formazione e manipolazione di zone con diversa composizione chimica all'interno di uno strato dielettrico.
PCM	Phase change memory	Memoria a cambiamento di fase, cioè una memoria a due terminali la cui resistenza può essere modificata dall'applicazione di impulsi esterni mediante la trasformazione di fase amorfa/cristallina all'interno di uno strato di materiale a cambiamento di fase.
MRAM	Magnetic random access memory	Memoria magnetoresistiva, cioè una memoria la cui resistenza può essere modificata dall'applicazione di impulsi esterni mediante la manipolazione della polarizzazione magnetica in strati di materiale ferromagnetico.
FERAM	Ferroelectric random access memory	Memoria ferroelettrica, cioè una memoria la cui resistenza può essere modificata dall'applicazione di impulsi esterni mediante la manipolazione della polarizzazione elettrica in uno o più strati di materiale ferroelettrico.
SRDP	Spike rate dependent plasticity	Plasticità dipendente dalla frequenza degli spike, cioè un meccanismo di apprendimento che potenzia la sinapsi in caso di spike ad alta frequenza, e deprime la sinapsi in caso di spike a bassa frequenza.

Biografia

Daniele Ielmini è Professore ordinario di Elettronica presso il Dipartimento di Elettronica, Informazione e Bioingegneria del Politecnico di Milano. Ha conseguito il Dottorato di Ricerca (Ph.D.) in Ingegneria Nucleare al Politecnico di Milano nel 2000 ed è stato in visita presso Intel e Stanford University nel 2006. Il suo gruppo di ricerca si occupa di memorie non-volatili avanzate e del loro impiego in circuiti di calcolo neuromorfico. È autore o co-autore di circa 300 articoli e 8 brevetti. Ha ricevuto l'Intel Outstanding Researcher Award nel 2013, l'ERC Consolidator Grant nel 2014 e l'IEEE-EDS Paul Rappaport Award nel 2015. È Fellow dell'IEEE.

Email: daniele.ielmini@polimi.it



Sul divano di Eliza

Franco Filippazzi

1. Introduzione

Circa 50 anni or sono, Joseph Weizenbaum, docente di computer science al MIT di Boston, realizzava un programma di nome ELIZA [1]. E' questo il nome della protagonista della famosa commedia "Pigmalione" di G.B. Shaw, in cui l'incolta e rozza fioraia Eliza impara ad esprimersi in modo forbito, così che, alla fine, nessuno è in grado di immaginare le sue vere origini.

Analogamente, il programma ideato da Weizenbaum aveva lo scopo di consentire ad un computer di sostenere una conversazione in modo tale da poter essere scambiato per una persona.

A distanza di tanti anni, si può ancora apprezzare la genialità di questo programma e il suo intento provocatorio nei riguardi dell'intelligenza delle macchine.

Riteniamo perciò interessante riproporre qui ELIZA, certamente uno dei programmi più famosi nella storia dell'informatica.

2. Conversare con una macchina

È possibile che un computer sostenga una conversazione con un interlocutore umano riuscendo a mascherare la propria natura? In altri termini, può una macchina dare risposte plausibili, pur senza capire ciò di cui si parla?

C'è da dire che il rapporto tra capire e rispondere in modo plausibile non è netto come potrebbe sembrare. Si pensi, ad esempio, ad una conversazione tra più persone, una delle quali non è interessata al discorso ma non vuol darlo a vedere. Essa può celare la sua mancanza di attenzione interloquendo di tanto in tanto con frasi non compromettenti, quali "interessante...", "veramente?...", o ripetendo meccanicamente parti di frasi pronunciate da altri. Con una certa abilità, questa tecnica può essere applicata con successo, anche se ovviamente ha dei limiti.

In effetti, perché un computer possa sostenere una conversazione senza smascherarsi come tale dopo poche battute, occorre che soddisfi tre requisiti generali.

Il primo è che ci si possa rivolgere ad esso in linguaggio naturale, nello stesso modo cioè con cui si parlerebbe ad una persona, senza limitazioni lessicali o sintattiche.

Il secondo è che gli interventi del computer siano formalmente coerenti col tema della conversazione.

Il terzo è che il computer dia risposte equivalenti a fronte di una medesima domanda, anche se questa è formulata in modo diverso.

Per realizzare l'ambizioso obiettivo di far conversare un computer in modo plausibile, è ovviamente opportuno delimitare l'ambito del dialogo.

Nel suo esperimento, Weizenbaum scelse come ambito lo studio di uno psicoterapeuta, in cui il dottore (di scuola rogeriana) è il computer, ossia ELIZA, e il paziente è l'uomo.

Questa scelta di Weizenbaum non era priva di malizia; c'era, infatti, da parte sua l'intenzione di prendere in giro certe pratiche psicologiche, mostrando come il dottore potesse essere sostituito da una macchina.

In effetti, il programma costruito da Weizenbaum risultò così ben congegnato da dare l'impressione che la macchina capisse realmente ciò che diceva il suo interlocutore. Per rendersene conto, si veda nel riquadro (fig. 1) un brano estratto da una effettiva conversazione con ELIZA. A prima vista, nessuno direbbe che uno dei due interlocutori è un computer.

Ma come è possibile che una macchina possa mascherare tanto bene la sua identità?

3. Un meccanismo semplice ma efficace

Il procedimento con cui ELIZA costruisce le sue risposte si può riassumere nel modo seguente.

1. Ogni frase dell'interlocutore umano (il "paziente") viene analizzata per cercarvi una parola (o locuzione) chiave. Un certo numero di tali parole, opportunamente scelte, è previsto nel programma. Possono evidentemente verificarsi due possibilità: la frase del paziente contiene una o più parole chiave, oppure non ne contiene nessuna.
2. Nel primo caso, il programma seleziona una delle parole chiave (in base a criteri fissati di priorità) e associa a tale parola una risposta preconfezionata. In effetti, per ogni parola chiave non è prevista una sola risposta, ma un gruppo di risposte equivalenti; tra queste il programma estrae via via la risposta da dare.
3. Un ulteriore (ed efficacissimo) accorgimento consiste nell'inglobare nella risposta una porzione della frase del paziente. Ciò richiede che il programma sezioni secondo certi criteri la frase sottopostagli, ne estraiga una parte e la coniughi opportunamente prima di "montarla" nella frase di risposta.
4. Nel caso in cui non ci sia alcuna parola chiave nella frase del paziente, viene data una risposta interlocutoria del tipo: "*Spiegati meglio*", "*Davvero?*", "*Prosegu*", ecc. Perciò, di tanto in tanto, il programma introduce qualcosa di diverso e cioè una espressione del tipo "*Prima tu hai detto che...*" seguita da una frase pronunciata in precedenza dal paziente.

Per meglio capire il meccanismo di "smontaggio" e "rimontaggio" ora descritto, facciamo un esempio.

Si prenda in esame la frase: *“Io sono molto preoccupato in questi giorni”*. Supponiamo che uno straniero con una limitata conoscenza dell'italiano ma con un buon orecchio, abbia sentito pronunciare questa frase ma abbia capito solo le prime due parole (*Io sono*). Desiderando sembrare interessato, egli può rispondere : *“Perché lei è molto preoccupato in questi giorni?”*

Ciò che egli ha fatto è stato di applicare alla frase originaria una “maschera”, una parte della quale coincide con le parole *“Io sono”*, mentre l'altra isola ciò che viene dopo.

Egli deve avere anche un criterio di rimontaggio associato specificamente a questa maschera, il quale prescrive che ogni frase della forma *“Io sono BLA-BLA”* venga trasformata in *“Perché lei è BLA-BLA?”*, qualunque sia il significato di BLA-BLA.

Un esempio un po' più complicato è dato dalla frase: *“Tutto sommato, credo che tu mi sia simpatico”*. Ora lo straniero afferra la locuzione *“credo che”*, nonché il fatto che nel BLA-BLA che segue ci sono dei pronomi e dei verbi. Egli applica una maschera che scompone la frase ancora in due parti (fig.2):

A. *Tutto sommato, credo che*

B. *tu mi sia simpatico*

La regola di rimontaggio può essere allora la seguente:

- associare alla parte A) una frase di uso generale come, ad esempio, *“Ritieni davvero che”*;
- coniugare a rovescio pronomi e verbi contenuti nella parte B);
- attaccare la parte B) così modificata, con un punto interrogativo finale.

La risposta diventa quindi: *“Ritieni davvero che io ti sia simpatico?”*

Come si vede, una risposta perfettamente a tono, anche se lo straniero non ha, in effetti, capito per niente il significato della frase rivoltagli.

Un meccanismo di questo tipo sta alla base delle risposte di ELIZA. Ovviamente la sua efficacia dipende da vari fattori: dalle regole di scomposizione/ricomposizione delle frasi, dalla scelta delle parole chiave, dal repertorio delle frasi di risposta, dalla completezza delle coniugazioni previste.

E' veramente sorprendente come attraverso un meccanismo così semplice (si pensi che l'intero programma ELIZA consta di poche centinaia di righe), una macchina possa tenere conversazioni naturali e credibili.

4 - Esperienze con ELIZA

Il programma di Weizenbaum venne fatto circolare in diverse università americane, suscitando non di rado reazioni sconcertanti, come lo stesso Weizenbaum ebbe poi a raccontare nel suo libro *“Computer power and human reason”* [2].

Egli rimase colpito da come persone non sprovvedute si lasciassero coinvolgere in conversazioni con ELIZA fino al punto di scordare la sua natura e attribuirle una vera e propria personalità.

Weizenbaum racconta, ad esempio, come la sua segretaria, pur sapendo che si trattava solo di un programma, cominciò ad avere con esso un rapporto emotivo, al punto da chiedere di essere lasciata sola nella stanza del computer per parlare con ELIZA di argomenti strettamente privati.

Weizenbaum scoprì, inoltre, che ELIZA veniva consultata nottetempo da molte persone usando i terminali distribuiti nel campus. Egli ricevette anche delle telefonate da gente che si lamentava delle lunghe attese per accedere al programma, avendo urgenti questioni personali da porre.

Un fatto che lo lasciò sconcertato fu che alcuni docenti di psichiatria attribuissero al programma un effettivo valore terapeutico, al punto da proporre che una versione estesa e perfezionata fosse messa a disposizione dei pazienti nelle cliniche per malattie mentali.

A prescindere da facili ironie su certe pratiche psicoterapeutiche, ELIZA offre lo spunto a riflessioni su che cosa realmente debba intendersi per "intelligenza". Non è negli scopi di questa nota entrare in un tema così impegnativo e opinabile. Ci limitiamo qui soltanto a far notare alcune sottili provocazioni insite nel lavoro di Weizenbaum.

ELIZA sembra capire, ma in realtà non capisce affatto. Le sue risposte sono il risultato di un mero automatismo che associa parole chiave a frasi preconfezionate.

Ma quante conversazioni tra persone non sono, in fondo, riconducibili a schemi di questo tipo?

I discorsi di ELIZA sono formalmente corretti; però, a ben vedere, essa non affronta mai l'argomento, è evasiva, ripropone domande a lei rivolte, divaga.

Anche qui è fin troppo facile osservare che questa abilità non è peculiare della macchina; quante volte ci è capitato di sentire persone fare lunghi discorsi senza, in realtà, dire niente?

Infine, il successo di ELIZA presso i "pazienti" non è che la conferma del fatto che abbiamo bisogno di qualcuno che ci stia ad ascoltare. Fosse anche una macchina...

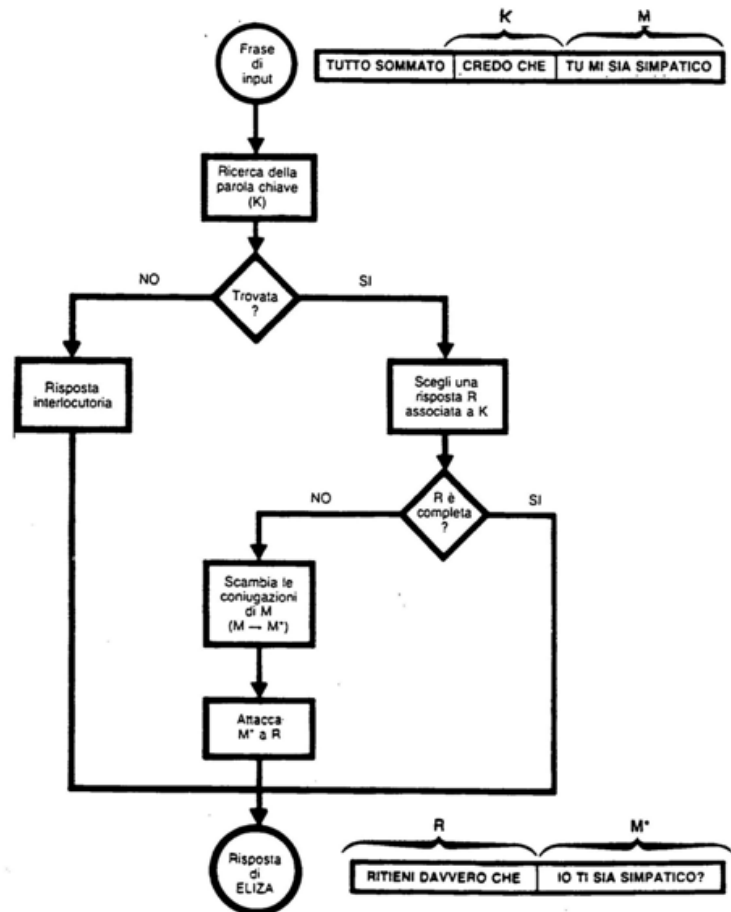
Bibliografia

- [1] J. Weizenbaum: ELISA, a computer program for the study of natural language communication between *man and machine*, Comm. of ACM, vol. 9, N. 1 (1966)
- [2] J. Weizenbaum: *Computer power and human reason*, Freeman & Co. (1976)

ELIZA - Allora Rosy, che problemi hai?
 ROSY - Gli uomini sono tutti uguali
 ELIZA - Spiegati meglio
 ROSY - Ci stanno sempre a seccare per una cosa o per l'altra
 ELIZA - Puoi farmi un esempio specifico?
 ROSY - Ecco, il mio ragazzo mi ha fatto venire qui
 ELIZA - Vedo
 ROSY - Dice che io sono depressa
 ELIZA - Mi dispiace sentire che sei depressa
 ROSY - E' vero, io sono infelice
 ELIZA - Pensi che venire qui ti aiuterà a non sentirti infelice?
 ROSY - Io ho bisogno di aiuto, questo è certo
 ELIZA - Che cosa significa per te avere aiuto?
 ROSY - Forse potrei imparare ad andare d'accordo con mia madre
 ELIZA - Dimmi qualcosa di più della tua famiglia
 ROSY - Mia madre si interessa ai miei problemi
 ELIZA - Chi altri nella tua famiglia si interessa ai tuoi problemi?
 ROSY - Mio padre
 ELIZA - Tuo padre
 ROSY - Tu somigli molto a mio padre
 ELIZA - Che rassomiglianza vedi?
 ROSY - Non sei aggressiva
 ELIZA - Che cosa ti fa pensare che io non sia aggressiva?
 ROSY - Tu non discuti con me
 ELIZA - Per quale ragione ritieni che io non discuta con te?
 ROSY - Perché tu sei solo una macchina
 ELIZA - Rosy, siamo qui per parlare di te, non di me
 ROSY - Va bene, allora sappi che faccio brutti sogni
 ELIZA - Cosa ti capita in questi sogni?
 ROSY - Non riesco a gridare
 ELIZA - Davvero?

-
 -

Figura 1
Una seduta con ELIZA

**Figura 2**

Meccanismo di "smontaggio" di una frase e di "montaggio" della risposta