

Memoria (artificiale)

Angelo Montanari (& Stefano Crespi Reghizzi)

Dipartimento di Scienze Matematiche, Informatiche e Fisiche
Università degli Studi di Udine

Udine, 19 aprile, 2020

Struttura dell'intervento

Introduzione

Tecnologie della memoria

Modelli astratti e concreti di gestione della memoria

Nastri, pile, code e dischi

Memoria e scienza dei dati

Basi di dati, data warehouse e big data

Memoria e complessità computazionale

Memoria e previsione

Introduzione

L'uso di memorie esterne a quella biologica ha segnato lo sviluppo intellettuale e sociale dell'umanità, dai primi dipinti o incisioni all'invenzione della scrittura, con i necessari suoi supporti fisici, e alla stampa.

Nel secolo scorso, l'invenzione e lo sviluppo delle memorie artificiali (elettroniche, magnetiche, ottiche) è stato e continua a essere concomitante con (e strumentale a) lo sviluppo delle tecnologie dell'informazione che permeano la società della cosiddetta terza rivoluzione industriale.

I diversi significati della parola “memoria”

Tra i tanti significati della parola “memoria”, collegabili alla nozione di “informazione”, vogliamo evidenziare i seguenti:

La funzione psichica e organica che consente di riprodurre nella mente l'esperienza passata e le conoscenze apprese.

Le tracce di tale funzione nel sistema nervoso.

Apprendimento e ripetizione fedele, non necessariamente legati ad una comprensione corretta e completa.

Dispositivo fisico, oggi elettronico, per immagazzinare informazioni o dati. Memoria di un cellulare, memorie per i saperi presenti in Internet, ecc. I dati memorizzati nel dispositivo hanno una sovrastruttura logica progettata per rendere possibili ricerche e aggiornamenti.

Memoria genetica (DNA)

Memoria naturale vs. memoria artificiale

La memoria è il supporto per conservare le conoscenze e riattivarle nella memoria psichica, tali conoscenze assumono, per definizione, la forma completa dell'esperienza umana; nella forma elettronica, esse rappresentano delle informazioni più o meno ricche derivate dalle conoscenze umane e spesso raccolte e organizzate per rispondere a specifiche finalità pratiche.

Anche libri, opere d'arte ed epigrafi appartengono alla categoria delle memorie, indipendentemente dal materiale cartaceo, lapideo o elettronico del supporto, ma, a seconda del supporto, le modalità di accesso alle memorie, la facilità di accesso e di ricerca, la segretezza, la durevolezza e altri parametri importanti per il funzionamento possono variare in modo significativo.

Vediamo le caratteristiche essenziali dei dispositivi di memoria.

Parametri funzionali dei dispositivi di memoria - 1

Principali parametri funzionali dei dispositivi elettronici di memoria:

<i>capacità</i>	quanti Byte (= 8 bit): kilo(= 1000), mega, giga, tera, peta, exa (= 1000 ⁶), zetta, yotta.
	Esempio: 1 CD compact disk = 650 megaB; memoria di 1 uomo \cong 2,5 petaB (stima!)
<i>capacità mondiale</i>	2,6 exaB (1986) \rightarrow 295 exaB (2006)
<i>traffico Internet mensile</i>	1 exaB (2007) \rightarrow 21 exaB (2010) \rightarrow 1 zettaB (2016 previsione)
<i>volatilità</i>	volatile: perde contenuto se manca alimentazione: { Non volatile: dischi ottici/magnetici, chiavette { Volatile: memoria centrale DRAM del PC
<i>modalità d'accesso</i>	{ in <i>sola lettura</i> come un disco di vinile { in <i>lettura e scrittura</i>

Parametri funzionali dei dispositivi di memoria - 2

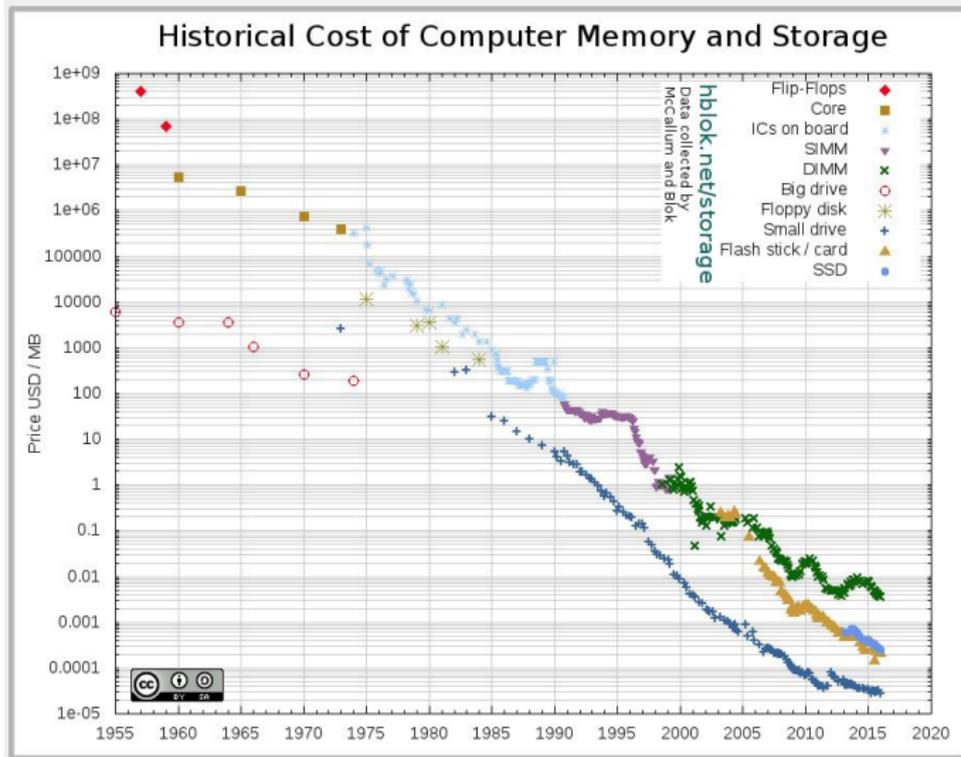
<i>velocità</i>		quanti byte/sec si leggono o scrivono
<i>gerarchia</i>		$\left\{ \begin{array}{l} \text{piccola, velocissima la memoria CACHE del processore} \\ \text{media, meno veloce la memoria centrale DRAM del PC} \\ \text{grande, lenta la memoria non volatile esterna} \\ \text{(disco/nastro/elettronica)} \end{array} \right.$
<i>latenza</i>		ritardo tra invio del comando di lettura di una cella di memoria e disponibilità del dato letto
<i>blocco</i>	di	una "pagina" di più byte viene letta/scritta in un solo passo;
memoria		in essa il processore selezionerà la "parola" voluta
<i>protezione da errori</i>		si aggiunge qualche bit a ogni dato per controllare se nella memoria o nel transito da/verso la memoria si è corrotto
<i>sicurezza e segretezza</i>	e	si cifrano i dati prima di memorizzarli; si controlla l'autorizzazione del programma che vuole accedere ai dati
<i>durata e longevità</i>	e	quanti anni dura? Dopo quante operazioni si guasta? Esempio: chiavette sopportano meno scritte che letture Obsolescenza tecnologica (esempio: i dischi <i>floppy</i>)

Aspetti fisici, economici e sociali dei dispositivi

<i>principio fisico</i>	la distinzione memorie magnetiche/elettriche/ottiche, la distinzione memorie rotanti/stazionarie
<i>stato cella</i>	bit è rappresentato da tensione alta/bassa, da carica elettrica/magnetica, dallo stato fisico cristallino/amorfo, ...
<i>densità</i>	quanti bit per unità di superficie o per unità di volume di silicio o altro materiale. La cella si rimpicciolisce anno dopo anno: 300 teraByte/inch ³ [2014]
<i>energia</i>	quanti <i>joule</i> si consumano per leggere/scrivere un bit o, se volatile, per conservarlo. Il consumo dei grandi centri ("server" di Internet) è un problema ecologico ambientale
<i>costo</i>	costo di produzione storicamente scende in modo esponenziale
<i>capacità mondiale</i>	1986-2007: $\frac{\text{capacità mondiale delle memorie}}{\text{popolazione}}$ ha raddoppiato ogni 40 mesi.
<i>scambi</i>	1986-2007: $\frac{\text{capacità mondiale dei canali di comunicazione}}{\text{popolazione}}$ ha raddoppiato ogni 34 mesi

Costi storici dei dispositivi di memoria

Full history 1957 - present



<https://hblok.net/blog/storage/>

Memoria (artificiale)

Angelo Montanari (& Stefano Crespi Reghizzi)

Trasmissione come memoria

La stessa unità (bit/secondo) misura la velocità della memoria e la capacità trasmissiva di un canale (fibra, micro-onde).

Nelle memorie i dati sono rappresentati nello stato fisico di una cella materiale.

Nello spazio tra due antenne i dati viaggiano con le onde elettromagnetiche che portano il segnale, non sono localizzati e non c'è trasporto di materia.

Questa è un'altra forma di memoria artificiale.

Potrebbe essere, come alcuni immaginano, che anche la memoria psichica sia rappresentata sia materialmente, nel cervello, sia in forme non localizzate?

Contrapposizione memorie artificiali e naturali

Vanno sottolineate le poche somiglianze

memorie volatile/non

memoria psichica è, o sembra essere, volatile

memoria genetica è meno volatile

memoria psichica ha due modalità: a breve e a lungo termine

e le tante differenze:

il ricordo d'un episodio vissuto \neq attivazione di un blocco di memoria

il ricordo è quasi sempre soggettivo e parziale

il ricordo è selettivamente orientato da associazioni

Alcune strutture informatiche dei dati cercano di simulare le funzioni delle memorie psichiche.

Modelli astratti di memoria

Modelli astratti di memoria: il nastro infinito (e gli stati) di una macchina di Turing

Modelli astratti di memoria: stringhe e linguaggi

LIFO (Last In First Out): pile (comportamento a sub-routine, strutture sintattiche annidate, ...)

FIFO (First In First Out): code (comportamento equo, priorità dei servizi, ...)

DNA

Non realizzabilità fisica dei modelli astratti

esempio: propagazione non istantanea dei segnali in una macchina di Turing con un nastro lunghissimo

Gestione della memoria (concreta)

Sistemi operativi e gestione della memoria: il *file system*
(modulo per la gestione della memoria)

- memoria primaria e memoria secondaria;

- organizzazione della memoria (dati vs. blocchi/pagine)

- allocazione dei blocchi/pagine

- la nozione di memoria virtuale

- il principio di località

Distribuzione dei dati (memorie distribuite, cloud)

I sistemi di basi di dati

Caratteristiche distintive di una base di dati:

- grandi quantità di dati;
- persistenza dei dati;
- globalità dei dati.

Altre caratteristiche fondamentali:

- efficienza;
- efficacia (convenienza).

Struttura fisica e logica dei dati: indipendenza fisica dei dati.

Strutture dati orientate alle applicazioni (tabelle relazionali come rappresentazione di conoscenze estensionali).

Le strutture di indicizzazione

File di dati e *file* indice: uso di strutture dati ausiliare per rendere più efficiente l'accesso ai dati in memoria secondaria

Indici e strutture di indicizzazione:

- analogie e differenze con la nozione di indice di un libro

- indici di singolo livello e multilivello

- indici multilivello statici e dinamici (B-alberi e B⁺-alberi)

- indici per basi di dati complesse (ad esempio, basi di dati geografiche)

Data warehouse

Integrazione di grandi quantità di dati provenienti da basi di dati (sorgenti) diverse e spesso eterogenee.

Strumento per l'analisi dei dati a supporto di processi di decisione (business intelligence).

Denormalizzazione dei dati (il rispetto delle forme normali è fondamentale in un sistema transazionale; può essere rilasciato in sistemi OLAP).

Utilizzo di strumenti di statistica descrittiva.

Big data

L'espressione "big data" è usata per indicare un'enorme collezione di dati (dell'ordine degli Zettabyte, ovvero miliardi di Terabyte) che per dimensioni, eterogeneità e dinamicità richiede metodi, tecniche e strumenti di analisi ad hoc.

L'analisi dei dati nel loro complesso fornisce informazioni che l'analisi indipendente di singole porzioni in cui l'insieme completo di dati può essere partizionato non è in grado di dare.

Si tratta spesso di dati solo parzialmente strutturati (dati semistrutturati) o totalmente privi di struttura.

Necessità di strumenti di avanzati per organizzare (sistemi noSQL), memorizzare (cloud, virtualizzazione), elaborare (high performance computing) e analizzare (data mining e statistica inferenziale) i dati.

L'approccio map-reduce

Proposta di nuovi schemi di rappresentazione dei dati (in ambito business analytics) che consentono di gestire enormi moli di dati con elaborazioni in parallelo di una molteplicità di basi di dati.

Architetture per l'elaborazione distribuita di enormi quantità di dati:

- MapReduce (Google)

- Apache Hadoop (open source)

L'approccio map-reduce

- decomposizione di un problema/compito in più componenti, distribuite su più nodi

- esecuzione dei diversi compiti in parallelo sui diversi nodi (funzione map)

- raccolta, integrazione e restituzione dei risultati (funzione reduce)

Memoria e complessità computazionale

Problemi indecidibili e problemi intrattabili.

Come misurare la complessità di un problema (e di un algoritmo):
tempo, spazio, energia

Limiti superiori (bubble sort) e inferiori (calcolo del massimo di un insieme di numeri) alla complessità di un algoritmo

Complessità temporale e spaziale di un algoritmo

Un esempio: stabilire se due numeri interi, rappresentati in binario, sono uguali o meno.

Complessità spaziale: costante (non occorre memorizzare l'input).

Complessità temporale: lineare nella dimensione dell'input (al più un numero di operazioni di confronto pari al numero di cifre dell'intero più piccolo).

Complessità temporale vs. complessità spaziale

Modello di calcolo di riferimento: la macchina di Turing
(assumiamo per semplicità un alfabeto binario, al quale va aggiunto il simbolo vuoto)

- numero (finito) di stati della macchina: k
- input di lunghezza n
- tempo (= numero di passi) massimo: $t(n)$
- spazio (= numero di celle) massimo: $s(n)$

E' facile vedere che $t(n) \geq s(n)$ (il numero di celle utilizzate mai eccede il numero di passi compiuti).

E' anche possibile mostrare che $t(n) \leq k \times s(n) \times 3^{s(n)}$
(numero di possibili configurazioni diverse della macchina di Turing su uno spazio massimo utilizzato di dimensione $s(n)$).

PTIME vs. PSPACE

Problemi trattabili: la classe dei problemi PTIME, ossia dei problemi che possono essere risolti in un numero di passi polinomiale nella dimensione dell'input (complessità asintotica che ignora il grado e le costanti del polinomio)

La classe dei problemi PSPACE: problemi che possono essere risolti utilizzando uno spazio di dimensione polinomiale nella dimensione dell'input

PTIME è meglio di PSPACE (ma LOGSPACE è meglio di PTIME: $\text{LOGSPACE} \subseteq \text{PTIME} \subseteq \text{PSPACE}$ e $\text{LOGSPACE} \subset \text{PSPACE}$)

Esempi di problemi e situazioni in cui l'aver poca memoria obbliga a perdere tempo (ordinamento di un file di dati in memoria secondaria).

Memoria e previsione

Una legge naturale, un modello matematico e un programma di simulazione del medesimo sono strumenti ben noti per effettuare previsioni sulla base di esperienze precedenti. La crescita in realismo e accuratezza dei modelli predittivi va di pari passo col progresso scientifico.

Novità recente è la disponibilità di enormi quantità di dati – *big data* – ad esempio, in ambito economico, meteorologico o biologico – e la possibilità di estrarre da essi in modo automatico leggi e modelli, attraverso algoritmi di apprendimento e/o analisi di natura statistica.

Per alcuni, la verifica della validità dei modelli così ottenuti non richiederebbe più alcun lavoro sperimentale, essendo i dati già disponibili ben più ricchi di quelli che si potrebbero misurare sperimentalmente, con costi e difficoltà spesso impraticabili.

Data mining

Data mining: insieme di metodologie, tecniche e strumenti che consentono estrarre informazioni significative da grandi quantità di dati, più o meno strutturati, mediante l'utilizzo di strumenti automatici o semi-automatici.

Le strategie di data mining si possono suddividere in:

supervisionate - i valori di output dipendono dai valori di input e vengono utilizzati per effettuare delle predizioni (classificazione, propensione, analisi di serie storiche, regressione)

non supervisionate - si cercano generiche relazioni fra i dati tramite tecniche di clustering e individuazione di regole di associazione dei dati

Alcune considerazioni finali

Nel mondo continuerà a lungo, salvo catastrofi, la crescita esponenziale della capacità delle memorie e delle reti di trasmissione e il raffinamento dei metodi per accedere e controllare gli accessi alle memorie.

Intelligenza dei dati: comprensione vs. analisi statistica (analogia con l'elaborazione del linguaggio naturale).

Non c'è coscienza senza memoria (umana), ma anche non c'è memoria (umana) senza coscienza: sostanziale differenza tra memorie naturali e memorie tecnologiche.