Memoria e previsione
Il punto di vista dell'informatica

Stefano Crespi Reghizzi e **Angelo Montanari**
Politecnico di Milano e Università degli Studi di Udine

Parole della Scienza
Roma, 21 ottobre, 2016

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
- Informazione e memoria
  - Teorie dell'informazione, teoria di Shannon, complessità di Kolmogorov
- 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:

capacità	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!)
capacità mondiale	2,6 exaB (1986) $\rightarrow$ 295 exaB (2006)
traffico Internet mensile	1 exaB (2007)$\rightarrow$ 21 exaB (2010)$\rightarrow$ 1 zettaB (2016 previsione)
volatilità	volatile: perde contenuto se manca alimentazione: { Non volatile: dischi ottici/magnetici, chiavette { Volatile: memoria centrale DRAM del PC
modalità d'accesso	{ in *sola lettura* come un disco di vinile { in *lettura e scrittura*

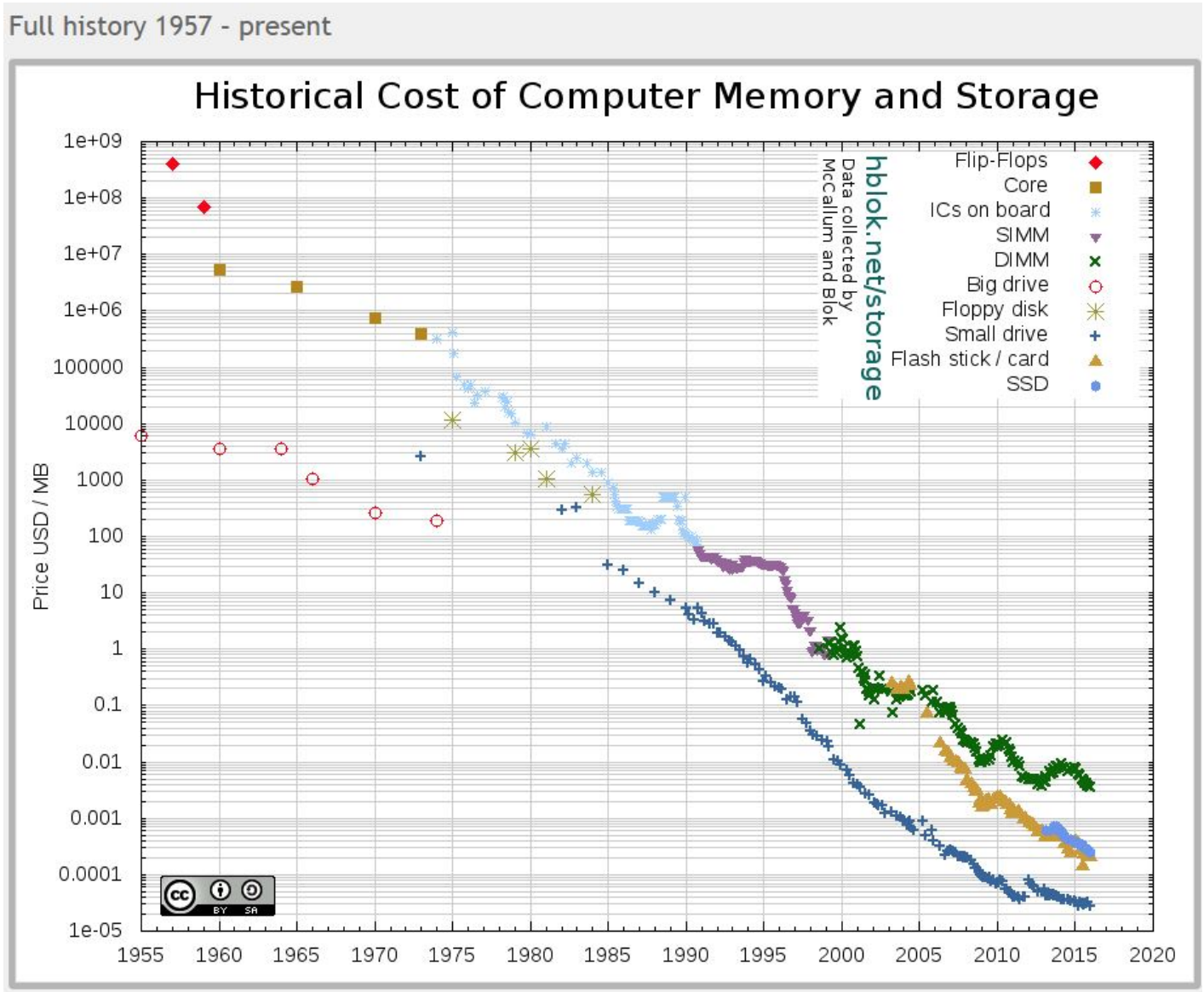
Parametri funzionali dei dispositivi di memoria - 2

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

Aspetti fisici, economici e sociali dei dispositivi

principio fisico	la distinzione memorie magnetiche/elettriche/ottiche, la distinzione memorie rotanti/stazionarie
stato cella	bit è rappresentato da tensione alta/bassa, da carica elettrica/magnetica, dallo stato fisico cristallino/amorfo, ...
densità	quanti bit per unità di superficie o per unità di volume di silicio o altro materiale. La cella si rimpicciolisce anno dopo anno: 300 $\mathrm{teraByte}/\mathrm{inch}^3$ [2014]
energia	quanti *joule* 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
costo	costo di produzione storicamente scende in modo esponenziale
capacità mondiale	1986-2007: $\frac{\text{capacità mondiale delle memorie}}{\text{popolazione}}$ ha raddoppiato ogni 40 mesi.
scambi	1986-2007: $\frac{\text{capacità mondiale dei canali di comunicazione}}{\text{popolazione}}$ ha raddoppiato ogni 34 mesi

Costi storici dei dispositivi di memoria



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

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 efficente 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 singoli 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, distribute 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 $t(n) \geqslant s(n)$ (il numero di celle utilizzate mai eccede il numero di passi compiuti).
- E' anche possibile mostrare che $t(n) \leqslant k \times s(n) \times 3^{s(n)}$ (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: LOGSPACE $\subseteq$ PTIME $\subseteq$ PSPACE e LOGSPACE $\subset$ PSPACE)

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

Memoria e informazione

Esiste, infine, un legame molto stretto tra le nozioni di memoria e di informazione che risultano collegate da molteplici significati.

La voce "Information" [Pieter Adriaans] della Stanford Encyclopedia of Philosopy si apre con la seguente definizione: "Il termine informazione nel linguaggio comune è oggi prevalentemente usato come sostantivo non numerabile per denotare una qualsiasi quantità di dati, codici o testi memorizzati, inviati, ricevuti o manipolati in un qualsiasi mezzo fisico".

La storia del termine informazione e delle varie nozioni ad esso correlate è complessa e, secondo Adriaans, resta ancora, in massima parte, da scrivere.

Teorie dell'informazione

Il significato del termine informazione cambia a seconda della tradizione filosofica e del contesto culturale e prammatico in cui viene usato.

I primi sforzi rivolti a formalizzare il concetto di informazione si ebbero nella seconda metà del secolo scorso, motivati dallo sviluppo dei sistemi di comunicazione, dei calcolatori e dei loro algoritmi.

Si può essere d'accordo con quanti vedono nelle diverse teorie formali dell'informazione (spiccano Shannon e Kolmogorov) il comune sforzo di rendere misurabili alcune proprietà estensionali delle conoscenze umane.

Teorie formali: Shannon [1948]

Shannon definisce in termini probabilistici la *quantità di informazione* di un testo (più in generale, oggetto digitale), all'interno di un insieme (finito o non) di testi aventi una distribuzione probabilistica nota. Prima di comunicare, trasmettitore e ricevitore devono accordarsi sulla codifica usata.

Sia X una variabile casuale definita nel dominio (enumerabile) $\mathcal{X}$ con distribuzione probabilistica $P(X = x) = p_x$.
L'*entropia di Shannon* (numero reale) è così definita:

$$H(X) = \sum_{x \in \mathcal{X}} p_x \log 1/p_x$$

Essa misura:

- la quantità di informazione che l'osservatore/ricevitore ha ottenuto dopo aver ricevuto la comunicazione che la variabile X vale x
- la lunghezza media in bit $\bar{L}$ dei codici più compatti possibili

$$H(X) \leqslant \bar{L} \leqslant H(X) + 1$$

Lacune della teoria di Shannon [Gruenwaldt e Vitanyi]

Shannon avverte che "*i messaggi hanno un significato . . . ma gli aspetti semantici della comunicazione sono irrilevanti per il problema ingegneristico* [della trasmissione]"

"*Qual è l'informazione di questo libro?*"

- Ha senso vedere un libro come un elemento dell'insieme di tutti i possibili libri, magari con una distribuzione di probabilità non uniforme nota?
- Ha senso misurare l'*informazione ereditaria codificata nel DNA* considerando una particolare forma animale come un elemento dell'insieme di tutte le possibili forme animali, con una distribuzione di probabilità sovrapposta?
- Se la distribuzione è uniforme, l'entropia dice quanti bit sono necessari per contare tutti i casi, ma tace il numero di bit necessari per trasmettere *ciascuno* dei messaggi individuali.

Un esempio

Si consideri l'insieme delle stringhe binarie lunghe 9999999999999999 ossia 10^{16} bit:

$$\overbrace{0\ldots00}^{9999999999999999\text{ bit}}\quad,\quad\overbrace{0\ldots01}^{9999999999999999}\quad,\quad\ldots,\quad\overbrace{1\ldots11}^{9999999999999999}$$

Per Shannon servono 9999999999999999 bit per rappresentare ciascuna stringa.

La stringa di soli uni $1\ldots11$ si può, però, codificare molto più compattamente: essa è la ripetizione del bit 1 per un numero di volte che, rappresentato come numero binario, richiede circa 50 bit; in totale bastano circa 55 bit! Ciò funziona a patto che chi riceve la stringa conosca l'algoritmo per decodificarla.

La descrizione di questa e molte altre stringhe che presentano delle regolarità può essere fortemente compressa.

Teorie formali: Kolmogorov [1965] o Solomonoff [1964] o Chaitin [1969]

Anche la complessità di *Kolmogorov* di un testo (più in generale, di un oggetto) si fonda sulla descrizione del testo (dell'oggetto) mediante una stringa di bit, ma senza che se ne definisca la distribuzione probabilistica all'interno di un insieme di casi possibili.

Un testo può avere più descrizioni, ma una descrizione deve descrivere soltanto un oggetto. Fra tutte le descrizioni d'un oggetto, si prende la *lunghezza della descrizione più breve quale misura della complessità dell'oggetto*. Tale quantità non è calcolabile in generale.

La funzione che dalla descrizione ricostruisce l'oggetto deve poter essere calcolata mediante una macchina di Turing, ossia mediante un programma. Il tempo richiesto per il calcolo non viene preso in considerazione nella teoria di Kolmogorov (ma in sviluppi successivi finora poco concludenti, sì)

Anche i numeri reali e il continuo

Gli oggetti di cui si misura la complessità appartengono all'insieme dei numeri interi o, più in generale, a un insieme discreto numerabile, ma anche i numeri reali, e quindi gli spazi continui, possono essere indirettamente trattati nelle teorie di Shannon e di Kolmogorov.

Ad esempio, per Shannon un segnale musicale può essere reso numerico campionando il segnale acustico ad intervalli fissati e applicando poi una scala discreta a tali misure.

Per Kolmogorov, una funzione che calcola numeri reali viene approssimata, quando possibile, da una macchina di Turing capace di calcolarla con accuratezza specificata; essa è espressa da un numero intero, fornito come ulteriore input alla macchina.

Ciò che manca è il significato

Né l'una né l'altra delle teorie prende in considerazione il *significato* dell'oggetto rappresentato, né tanto meno se l'informazione rappresentata sia *corretta, coerente* o *vera*. Quindi tali formalizzazioni della nozione di informazione restano ben al di sotto, non solo della nozione di conoscenza umana, ma persino dei criteri di progetto delle basi-dati, applicati in informatica.

In uno studio recente, il logico J.M. Dunn (citato in Adriaans) dice bene che cosa sia l'informazione nel senso delle teorie formali prese in considerazione in precedenza: "*what is left of knowledge when one takes away believe, justification, and truth*".

Alcuni di questi aspetti sono considerati in altri studi teorici di natura logica e filosofica, ma troppo distanti dall'ambito scientifico oggetto del presente intervento per essere qui esaminati. Essi non sembrano aver avuto finora grande seguito [Adriaans].

Memoria e previsione: automi e giochi

- Memoria e previsione in un automa a stati finiti: le nozioni di stato e di funzione/relazione di transizione.
- Church e il problema della sintesi.
- Strategie posizionali (e non) di un gioco logico.

Memoria e previsione: inferenza induttiva e data mining

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.

Inferenza induttiva [Solomonoff]

Scegliere un'ipotesi (o legge) consistente con i dati fenomenici noti e capace di prevedere i prossimi è come *estrapolare da una serie* o stringa di numeri il valore del prossimo numero (vedi certi test psicologici di intelligenza).

Processo continuo di apprendimento: procede per falsificazione e correzione dell'ipotesi corrente alla luce di nuovi dati pervenuti. E' posto da Solomonoff in un contesto probabilistico bayesiano.

Le *ipotesi possibili* possono essere infinite ma *enumerabili*. L'ipotesi corrente è una *macchina di Turing* che deve accettare l'intera serie passata e può enumerare (= *prevedere*) i dati futuri.

Inferenza induttiva e criterio di Occam

Criterio di scelta dell'ipotesi conforme al *criterio di Occam*: la macchina di Turing più semplice, ossia descritta dal numero minimo di bit (Kolmogorov). Essa fornisce una descrizione compressa dei dati, sfruttando certe regolarità in essi presenti.

Apprendere corrisponderebbe allora a comprimere i dati grezzi.

Ma non è detto che in tal modo l'ipotesi scelta descriva davvero gli aspetti significativi presenti nei dati. Infatti la complessità di Kolmogorov ignora la distinzione tra diversi *generi* di informazione, quali l'informazione *significativa o interessante* e *non*.

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, da cui la limitatezza psicologica dei modelli cognitivi umani fondati sull'analogia computazionale, proposti dai pionieri dell'IA e entusiasticamente adottati da tanti psicologi.
- Non per questo l'IA perde di importanza: essa è una straordinaria amplificazione dell'intelligenza e dell'operatività umane in molti campi.