

Tutto_Misure, 2, 2026

[18.4.26]

Nella prima parte di questo articolo, pubblicata nel numero precedente di Tutto_Misure (1, 2026), abbiamo accennato all'interesse a comprendere *il ruolo della misurazione nella relazione tra dati e informazione*, e a questo scopo abbiamo introdotto una possibile distinzione appunto tra dati e informazione, interpretati, in accordo alla teoria dell'informazione di Claude Shannon, come *informazione sintattica e informazione semantica* rispettivamente.

A proposito di dati abbiamo già presentato e commentato i punti fondamentali del quadro concettuale: (i) si ottiene un dato / informazione sintattica dalla scelta di un'entità da un insieme di almeno due elementi (per esempio la scelta di una lettera da un alfabeto di almeno due lettere); (ii) meno è probabile la scelta, maggiore è la quantità di informazione sintattica ottenuta (in italiano la "h" porta più informazione sintattica della "a"); (iii) possono essere definite delle regole sulla struttura dei dati, che stabiliscono quali successioni di elementi dell'insieme sono "ben formate" (in una parola italiana dopo una "h" non ci può essere una consonante).

Quando l'insieme è discreto, Shannon usa il termine "simboli" per gli elementi dell'insieme (sono perciò simboli le lettere dell'alfabeto), e, con enfasi sulla comunicazione, il termine "messaggi" per le successioni di simboli. Potremmo allora dire che un simbolo – o la scelta di un simbolo: non è rilevante distinguere qui – è un dato elementare, e un messaggio è un dato strutturato, o composto.

Il termine "simbolo" è rivelatore qui di una questione che scopriremo essere fondamentale per la misurazione. Nel mondo empirico ci sono cose come pianeti e atomi, con le loro proprietà, ma non ci sono simboli (e analogamente, anticipiamo, ci sono oggetti e grandezze di oggetti da misurare, ma non ci sono valori di grandezza): nulla è in sé un simbolo, perché siamo noi a stabilire che certe cose in certi contesti sono simboli, per esempio quando impariamo a leggere e quindi a riconoscere che certe macchie di inchiostro sono il "supporto fisico" per certi simboli che sono certe lettere dell'alfabeto. Per una persona che non ha ancora imparato a leggere, quelle macchie di inchiostro sono solo entità fisiche con le loro proprietà fisiche. Leggere, come misurare, è dunque un atto che trascende il mondo fisico, perché produrre dati è già il risultato di un processo di astrazione. Quelle due macchie di inchiostro di fronte a noi non hanno nemmeno un atomo in comune eppure, disinteressandoci di molte loro proprietà fisiche, potremmo riconoscere entrambe come realizzazioni di uno stesso simbolo: potremmo cioè riconoscere che portano lo stesso dato, per esempio la lettera "a" (analogamente, due correnti elettriche potrebbero essere flussi di elettroni in circuiti diversi, e ciononostante avere la stessa intensità).

Non solo dati

In tutto ciò i dati possono non avere / portare alcun significato, e quindi il passaggio da dati a informazione – o dunque da informazione sintattica a informazione semantica – è ancora da compiere. Questa interpretazione è consistente con uno schema costruttivo bottom-up, una "piramide dati-informazione-conoscenza" che parte dai dati e tratta l'informazione come dati con semantica e la conoscenza come informazione con pragmatica.

(Per apprezzare come questa impostazione non sia così ovvia si può per esempio consultare la norma ISO/IEC 2382:2015 Information technology – Vocabulary, che segue una strategia opposta, top-down, definendo il concetto di dato in funzione di quello di informazione ("data: reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing") e il concetto di informazione in funzione di quello conoscenza ("information: <information processing> knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning" e anche "information: <information theory> knowledge which reduces or removes uncertainty about the occurrence of a specific event from a given set of possible events"), e infine

definendo ‘conoscenza’ (“knowledge: <artificial intelligence> collection of facts, events, beliefs, and rules, organized for systematic use”) in modo quantomeno opinabile.)

A proposito dell’interesse a costruire questa piramide, Shannon è esplicito (le citazioni che seguono sono mie traduzioni da *A mathematical theory of communication*, il famoso articolo pubblicato sul Bell System Technical Journal nel 1948): poiché “il problema fondamentale della comunicazione è di riprodurre in un punto esattamente o approssimativamente un messaggio scelto in un altro punto”, il fatto che “frequentemente i messaggi hanno un significato” è “irrilevante per il problema ingegneristico”. Come discuteremo, non è così per la misurazione, che ha la sua ragione nella produzione di significato a proposito del misurando e in questo si differenzia perciò in modo sostanziale dalla trasmissione. Dunque riprendiamo la costruzione della piramide.

Forma e contenuto

La strategia di astrazione è ancora la nostra guida. Se sistemi fisici diversi possono essere realizzazioni dello stesso simbolo o messaggio, e quindi portare lo stesso dato, l’intuizione basilare a proposito della relazione tra dati e informazione è analoga: sotto condizioni da chiarire, dati diversi possono portare la stessa informazione.

Vediamolo con un esempio. Supponiamo che l’insieme A dei dati elementari contenga i caratteri dell’alfabeto, “a”, “b”, ..., e che tutte le successioni di questi caratteri siano ben formate. Consideriamo ora un nuovo insieme B, ottenuto aggiungendo ad A i numeri interi positivi e le parentesi (“(” e “)”), con la condizione sintattica che un numero deve essere seguito da “(”, quindi uno o più caratteri, quindi “)”. Dunque per esempio “abcabcabcabc” è una successione ben formata sia per A sia per B, mentre “4(abc)d” è una successione ben formata per B ma non per A. D’altra parte, assumendo che simboli che sono cifre rappresentino effettivamente numeri (si noti che in effetti questa è già una prima semantizzazione, che introduciamo per mantenere l’esempio semplice) e adottando la regola per cui per esempio “2(ab)” in B corrisponde a 2 volte “ab”, cioè “abab”, in A, se ne conclude che le due successioni di dati “abcabcabcabc” e “4(abc)d” sono diverse ma, a meno di conoscere la regola, si possono far corrispondere l’una all’altra, in una relazione di equivalenza. Con forse una certa forzatura lessicale, potremmo dire che il loro *contenuto* è lo stesso. Poiché la seconda successione è più breve, in numero di caratteri, della prima, ne concludiamo allora che può essere possibile comprimere una successione di dati / un messaggio senza perdere nulla in contenuto (*lossless compression* in inglese). Siamo dunque di fronte a una distinzione operativa tra dati e contenuto, che si applica a situazioni di compressione di dati ma non solo, e che ci è ben familiare: per esempio, sotto ben note condizioni “1,23” e “1.23”, “metre” e “meter” hanno lo stesso contenuto. D’altra parte, l’esempio che abbiamo proposto sopra mostra che contenuto (per come lo abbiamo usato) e significato non coincidono: “ababab” e “3(ab)” hanno lo stesso contenuto, ma potrebbero non avere propriamente alcun significato.

In altri termini, un primo passo per trasformare dati in informazione è di riconoscere la *distinzione tra forma e contenuto*: se i dati sono la forma del supporto fisico su cui sono realizzati, ci può essere qualcosa oltre la forma. E infatti nelle trasformazioni mostrate sopra la forma cambia, e può essere eventualmente ridotta di dimensioni – in questo caso in numero di caratteri – senza che cambi il contenuto (sappiamo che il limite inferiore della riducibilità è stabilito dall’entropia di Shannon, ma l’argomento non è specificamente rilevante qui). Con la scoperta che dati diversi / forme diverse possono essere messi/e in corrispondenza, la costruzione della “piramide dati-informazione-conoscenza” prosegue, dunque.

Verso i significati

Nella prima parte di questo articolo, a proposito dei tre “livelli” di problemi della comunicazione abbiamo accennato all’idea di informazione con significato, e dunque di informazione semantica, proponendo la visione semplice secondo cui una semantica è una relazione tra due sintassi, una relativa a una struttura di dati (i significanti, per esempio le parole della lingua italiana) e l’altra relativa ad altre entità (i significati, nell’esempio tutto ciò che le parole della lingua italiana possono designare – incluse le parole stesse, dato che possiamo parlare in italiano della lingua italiana, come quando diciamo che “misurazione” è un sostantivo femminile singolare).

La relazione tra supporto fisico e dato è condizione necessaria ma non sufficiente per stabilire una relazione tra dato e informazione (dunque ripetiamo, per chiarezza: tra informazione sintattica e informazione semantica, nel senso di Shannon). Potremmo saper leggere la parola “misurazione”, e acquisire con ciò un dato, ma non sapere cosa essa significhi, cioè cosa sia una misurazione.

Nel suo libro *Syntactic Structures*, pubblicato nel 1957, Noam Chomsky ci dà un ottimo esempio di questa non sufficienza, con la frase “*Colorless green ideas sleep furiously.*”. È sintatticamente corretta benché sia davvero difficile attribuirle un qualsiasi significato, perché qualcosa non può essere contemporaneamente verde e senza colore e perché le idee non dormono e comunque non dormono furiosamente. È un dato, sì, che per esempio richiede un certo numero di bit per poter essere memorizzato su un supporto digitale, ma è problematico che sia anche informazione. Dal punto di vista del problema tecnico della comunicazione, nel senso di Shannon, la frase di Chomsky è solo una successione di un certo numero di simboli, ognuno con la sua probabilità di essere prodotto, e in questo non è diversa da una qualsiasi altra successione di lettere del nostro alfabeto. E il fatto che “cerchi di produrre significato” senza riuscirci mostra che per stabilire una relazione tra dati e informazione ci sono delle condizioni da soddisfare.

Noi vogliamo che un risultato di misura si riferisca al misurando e lo faccia in modo significativo. Dunque a proposito di misurazione la dimensione semantica è irrinunciabile, e le condizioni perché questa sia possibile, ed efficace, sono nella struttura stessa del processo di misurazione. Il ruolo di questa struttura e degli strumenti di misura come generatori di significato sarà l’argomento della prossima parte di questo articolo.