

La conoscenza in intelligenza artificiale

Stefano Mizzaro

Dipartimento di Matematica e Informatica, Università di Udine
V. delle Scienze, 206 – Loc. Rizzi – 33100 Udine – Italy
e-mail: mizzaro@dimi.uniud.it

Sommario: L'obiettivo di questo lavoro è indagare sul concetto di conoscenza al fine di ottenere indicazioni utili per le tecniche di intelligenza artificiale e rappresentazione della conoscenza. Alcune definizioni di conoscenza proposte da vari autori vengono analizzate, confrontate e classificate; esse vengono poi fuse in una nuova definizione, la quale viene parzialmente formalizzata con strumenti logico-algebrici; infine, vengono illustrati i possibili collegamenti con i settori dell'intelligenza artificiale e della rappresentazione della conoscenza.

1 Introduzione

Che cos'è la *conoscenza*? I filosofi hanno indagato per millenni su questo problema, senza giungere, per quanto ne so, ad una risposta definitiva. Non sarò quindi così presuntuoso da cercare di rispondere in modo completo a tale domanda: cercherò piuttosto di presentare alcune osservazioni e riflessioni, riguardanti la conoscenza, nella maniera più organica possibile.

L'obiettivo di questo studio è indagare sulla natura del concetto di conoscenza (da qui in poi K , dall'inglese *knowledge*) allo scopo di ottenere informazioni utili per le tecniche di rappresentazione della conoscenza (KR , da *knowledge representation*) o più in generale di intelligenza artificiale (IA). In queste note non espongo certezze consolidate, ma idee ed argomenti su cui sto tuttora riflettendo.

Si può indagare sulla K da differenti punti di vista:

- logico-formale: utilizzare gli strumenti della logica matematica per dare definizioni e dimostrazioni formali. Tale metodologia è quella i cui risultati sono senz'altro meno discutibili, ma allo stesso tempo probabilmente più limitati;
- psicologico-cognitivo: effettuare esperimenti psicologici su un campione di soggetti. La validità di questo metodo è assicurata da considerazioni statistiche, ma esso richiede uno sforzo non indifferente;
- filosofico-epistemologico- IA : riflettere, in modo intuitivo, su considerazioni di carattere generale. È senza dubbio il punto di vista meno "scientifico", ma quello secondo me più adatto alla fase iniziale di uno studio. È infatti questa l'ottica adottata qui.

Come accennato, la K è da molto tempo oggetto di studio dei filosofi, senza risultati definitivi. Negli ultimi decenni, grazie allo sviluppo del settore dell' IA , e più in particolare del settore della KR , il termine "conoscenza" è ritornato di moda, senza che per questo si sia giunti ad una completa comprensione del fenomeno. Si potrebbe ritenere infatti che i libri di IA e KR rispondano alla domanda iniziale di questo lavoro, ma tale aspettativa è ben presto delusa: i testi più diffusi iniziano le loro analisi direttamente dalla *rappresentazione* della K e non dalla K stessa. Nei (pochi) casi in cui non è così, solitamente gli autori se la cavano con affermazioni superficiali del tipo "la K è costituita da dati messi in relazione con oggetti del mondo", o qualcosa di analogo. Per convincersi di quanto sia trascurato lo studio della K nei testi di IA e KR , si vedano ad esempio: [1, 2, 5, 12, 15].

Altre due osservazioni, di tipo linguistico, che indicano con quanta superficialità sia spesso affrontato l'argomento nel settore dell' IA : in italiano, il termine inglese 'knowledge' andrebbe tradotto con 'conoscenze', e quindi 'knowledge representation' con 'rappresentazione delle conoscenze', 'knowledge base' con 'base di conoscenze', e così via. È invece abitudine comune parlare di 'conoscenza', di 'rappresentazione della conoscenza' e di 'basi di conoscenza', a parte alcune encomiabili eccezioni (ad esempio [4]). Considerazioni analoghe si possono effettuare per i termini 'conoscere' e 'sapere': l'inglese 'knowledge' viene infatti solitamente tradotto, nel settore della KR , con 'conoscenza' e 'to know' con 'conoscere'. Ma le alternative 'sapienza' (o 'il sapere') e 'sapere' mi sembrano altrettanto valide: si potrebbe dunque parlare di 'basi di sapienza' o di 'rappresentazione della sapienza', eppure tali espressioni non vengono utilizzate.

Questo lavoro è strutturato nel modo seguente: nel paragrafo 2 presento e discuto alcune definizioni di K ricavate dalla letteratura; nel paragrafo 3 riassumo le caratteristiche a mio avviso corrette in una ulteriore definizione; nel paragrafo 4 delinea una possibile formalizzazione della situazione; nel paragrafo 5 metto in relazione i risultati dell'indagine effettuata sulla K con l' IA e infine nel paragrafo 6 presento alcune osservazioni conclusive.

2 Conoscenza: alcune definizioni

In questo paragrafo sono riportate alcune definizioni (o pseudo-definizioni) di K reperite in letteratura, prevalentemente in lavori di IA , e rappresentanti le opinioni più diffuse. Ho suddiviso tali definizioni in cinque

categorie, denominate nel modo seguente:

- (i) K = dati astratti;
- (ii) K = dati e relazioni fra di essi;
- (iii) K = dati correlati al mondo reale;
- (iv) K = dati per agire;
- (v) K = proprietà per spiegare il comportamento.

Il significato di tali denominazioni risulterà chiaro nel seguito. Ognuno dei prossimi cinque sottoparagrafi contiene alcune definizioni di uno dei cinque gruppi.

2.1 Conoscenza = dati astratti

Le due citazioni seguenti sono rappresentative della categoria (i):

If we can trust an automatic process or clerk to collect the material then we are talking about *data*. [...] If we look for an expert to provide the material then we are talking about *knowledge*. Knowledge includes abstraction and generalization of voluminous data.

[16], p. 79

[Knowledge bases] include large amounts of abstract knowledge [...] and a small amount of (less interesting) concrete knowledge [...] [Data bases] were designed mostly for concrete knowledge, and the modelling methods were inadequate for representing abstract knowledge.

[18], p. 344

Secondo tali autori, la K si ottiene dai dati mediante operazioni di astrazione e generalizzazione; questa opinione è a mio avviso largamente diffusa nel campo delle basi di dati, o nel settore in cui si studia l'integrazione di basi di conoscenze e basi di dati o ancora nel settore di recente sviluppo del *data mining*, nel quale si ricercano, con metodi di apprendimento induttivo, relazioni e schemi globali esistenti in grandi basi di dati (si veda ad esempio [6]).

Altre definizioni di K che a mio parere potrebbero essere favorevolmente accolte in tali settori sono quelle appartenenti alla categoria (ii).

2.2 Conoscenza = dati + relazioni fra dati

Le definizioni della categoria (ii) potrebbero essere riassunte con: «i dati stanno alle conoscenze come gli insiemi stanno ai grafi». Più in dettaglio, si consideri un grafo $G = (N,A)$, in cui N è l'insieme dei nodi e A l'insieme degli archi rappresentanti relazioni fra tali nodi. Se $A = \emptyset$, e quindi se non si hanno relazioni fra i nodi (o quantomeno se si ha un numero limitato di tali relazioni), si ottiene una base di dati; arricchendo l'insieme A delle relazioni fra nodi ci si sposta verso una base di conoscenze.

Questo gruppo di definizioni può essere ulteriormente scomposto: infatti si può decidere di attribuire a G lo status di basi di dati unicamente se $A = \emptyset$, oppure si può essere meno drastici e considerare basi di dati anche casi in cui A abbia un numero limitato di tali relazioni. Nel primo caso si ha una distinzione di tipo *qualitativo* fra dato e K, mentre nel secondo caso la differenza fra dato e K risulta molto sfumata ed è unicamente di tipo *quantitativo*: dati e K si vengono a collocare ai due estremi di un *continuum* e non vi è una linea di separazione netta fra di essi. Questa seconda posizione è ispirata alla distinzione fra sintassi e semantica proposta da G. O. Longo:

Searle dà per scontato che la semantica sia una proprietà o capacità "tutto o niente", che o si possiede o non si possiede. Ma è proprio vero? La semantica non è invece forse una proprietà sfumata, che si può avere in vari gradi? La semantica, forse, è una sorta di corrispondenza estesa, di pseudo isomorfismo, o di quasi isomorfismo (il nome non ha molta importanza) dinamico, cioè in evoluzione, che ogni essere umano si costruisce e mantiene tra le diverse mappe mentali e corporee, tra le diverse sintassi, o grammatiche, create e gestite dai suoi sensi e dalla sua conoscenza razionale.

È questa fusione, o quasi fusione, dinamica di più sintassi (e della rappresentazione, di continuo aggiornata e controllata, di questa fusione di più sintassi) che costituisce la semantica.

[7], p. 18

Se infatti si estrapolano queste considerazioni a dati e K, sostituendo a sintassi e semantica, rispettivamente, dati e K, si ritrova la posizione precedentemente indicata con (ii).¹ Si osservi che Longo attribuisce alla semantica un carattere «dinamico, cioè in evoluzione». Tale dinamicità, se si accetta l'estrapolazione che ho appena proposto, dovrà essere posseduta anche dalla conoscenza.

¹ In realtà, il parallelo fra la proposta di Longo e la distinzione fra dati e K non è esatto, in quanto la distinzione di Longo è più sottile. Per Longo, infatti, ogni sintassi costituisce un grafo e la semantica si ottiene ponendo in relazione elementi dei vari grafi delle sintassi; nel caso delle definizioni del gruppo (ii), invece, vi è un unico grafo.

2.3 Conoscenza = dati correlati al mondo reale

Le definizioni del gruppo (iii) sono forse le più diffuse nel settore logicista dell'IA. Le prime parole della prefazione di [15] sono:

Knowledge-based systems emphasize meaning. Instead of processing data as a string of bits, they represent the meaning of data in terms of the real world.

[15], p. vii

Probabilmente, anche il filosofo J. Searle sarebbe d'accordo con questa distinzione. Egli infatti, seppure in un contesto differente, afferma che:

[...] gli atti di pensiero, di percezione, di comprensione e così via hanno un contenuto mentale. Grazie al loro contenuto possono concernere oggetti e situazioni del mondo esterno.

[13], p. 17

Il problema delle definizioni del gruppo (iii), e delle posizioni *realiste* in genere, è che esse fanno riferimento al *mondo reale*, ente che però non può essere pienamente compreso: la percezione che noi abbiamo del mondo reale è infatti filtrata dai nostri sensi e dalla nostra coscienza; inoltre la *K* può riguardare concetti astratti, slegati dal mondo reale.

Si rifletta un attimo sulle distinzioni fra sintassi e semantica che si incontrano nei settori della logica matematica, della semantica formale dei linguaggi di programmazione e della semantica del linguaggio naturale. Tali distinzioni sono analoghe a quella fra dati e *K* proposta qui e ne sono sicuramente state ispiratrici. Al contrario, le definizioni del gruppo (ii) possono essere paragonate alla semantica algebrica.

2.4 Conoscenza = dati per agire

Anche le definizioni appartenenti alla categoria (iv) sono piuttosto diffuse. Ad esempio, Newell afferma:

If a system has (and can use) a data structure which can be said to represent something (an object, a procedure, ... whatever), then the system itself can be said to have knowledge, namely the knowledge embodied in the representation about that thing.

[11], p. 90, il corsivo è mio

Può sembrare che la definizione di Newell ricada nella categoria (iii), ma non è così in quanto contiene un aspetto ulteriore (invero espresso piuttosto marginalmente): *la K va usata*.

Sempre sulla stessa linea è Brian C. Smith, che esprime così la sua *ipotesi della rappresentazione delle conoscenze*:

Any mechanically embodied intelligent process will be comprised of structural ingredients that a) we as external observer naturally take to represent a propositional account of the knowledge that the overall process exhibits, and b) independent of such external semantical attribution, play a formal but causal and essential role in engendering the behaviour that manifests that knowledge.

[14], p. 2

Nelle ultime due citazioni sono illustrati, più o meno esplicitamente, due punti a mio avviso essenziali:

- la *K* va usata per agire, ed è quindi da collegare con il *comportamento*, con le *azioni* dell'ente che la possiede;
- le strutture dati interne di un programma, che sono viste come dati dall'interprete che esegue tale programma, dall'esterno possono essere viste come conoscenze.

Ritorrerò in seguito sui collegamenti fra *K* e azione. Apro invece una parentesi sul secondo dei punti precedenti: lo stesso oggetto può essere visto in modi differenti, come strutture dati o come *K*. Anche in [8] si trova una posizione analoga:

Immaginiamo un soggetto che è vissuto per tutta la vita in un sottomarino e che, non essendone mai uscito, è perfettamente addestrato a guidarlo. Ora noi siamo sulla spiaggia e vediamo che il sottomarino si avvicina ed emerge dolcemente alla superficie; prendiamo la radio e diciamo al pilota che sta all'interno: «Congratulazioni, hai evitato gli scogli e sei emerso con grande eleganza; le manovre del sottomarino sono state perfette». Il nostro pilota all'interno, però, è sconcertato: «Cosa è questa storia di scogli e di emersione? Tutto quello che ho fatto è stato muovere leve e girare manopole e stabilire certe relazioni fra indicatori in una sequenza preordinata in accordo con le mie abitudini. Io non ho effettuato alcuna manovra e tu mi parli di un sottomarino, per di più: mi sembra quasi una burla».

Per l'uomo all'interno del sottomarino esistono solo le letture degli indicatori, le loro transizioni e i modi di ottenere certe relazioni specifiche tra esse. Solo per noi, che stiamo fuori e vediamo come cambiano le relazioni fra il sottomarino e il suo ambiente, esiste il comportamento del sottomarino [...]

non dobbiamo confondere il funzionamento del sottomarino, la sua dinamica di stati, con i suoi spostamenti e cambiamenti di posizione nell'ambiente.

[8], p. 118

Vi sono quindi due livelli, o domini, o punti di vista, differenti: l'agente che mantiene una sua dinamica interna ("data structure" nella citazione precedente di Newell e punto *b*) della citazione di Smith) e l'osservatore che vede l'agente interagire con il suo ambiente ("knowledge embodied in the representation" in Newell e punto *a*) in Smith). E questi due livelli, secondo Maturana e Varela, vanno tenuti separati.

In questa sede non indago ulteriormente tali problematiche; mi limito a formulare alcune domande che ritengo indicative della complessità del problema. Innanzi tutto: quanto sono realmente osservabili i due livelli? Essi vanno davvero tenuti separati? Se sì, come si fa a spiegare il comportamento di un'agente? Se l'agente e l'osservatore coincidono (e questo accade comunemente per noi esseri umani: tramite auto-osservazione, l'uomo riesce a "vedere" entrambi i livelli), tale separazione è ancora possibile?

Inoltre, come interviene in questo quadro il fenomeno della *coscienza*? Non è forse questa una caratteristica indispensabile per la K? In che misura l'aver K facilita l'adattamento di un ente al suo ambiente? Che collegamenti ci sono con l'apprendimento? L'attitudine tipicamente umana di "raccontarsi storie" è una conseguenza o una causa della K (o della coscienza, o dell'adattamento all'ambiente)? Che cosa è che differenzia la K posseduta dagli esseri umani da quella posseduta dagli animali?

Con questa carrellata di domande, che danno un'idea di quanta strada sia ancora da compiere prima di comprendere interamente la situazione, chiudo la parentesi sui due livelli della dinamica interna dell'agente e del comportamento nel suo ambiente. Nel prossimo paragrafo presento l'ultimo gruppo di definizioni di K tratte dalla letteratura.

2.5 Conoscenza = proprietà per spiegare il comportamento

Tutte le definizioni riportate fino qui hanno in comune il fatto di basarsi sul concetto di *dato* e di arricchirlo in qualche modo per ottenere il concetto di conoscenza. Questa strada non sembra portare a risultati definitivi, forse a causa del fatto che neppure il concetto di dato è pienamente compreso, o forse perché la distinzione fra dato e K è unicamente di tipo quantitativo (e non qualitativo) e quindi non ha senso arricchire in modo qualitativo il concetto di dato per arrivare a quello di K. Per questi motivi, le definizioni precedenti sono, secondo me, non solo incomplete, ma fuorvianti; le prossime definizioni, appartenenti alla categoria (v), vanno in una direzione differente e cercano di definire la K indipendentemente dal dato.

La prima definizione riportata qui e appartenente al gruppo (v) è di Genesereth e Nilsson:

Intelligent entities seem to anticipate their environments and the consequences of their actions. They act as if they know, in some sense, what the result would be. We can account for this anticipatory behavior by assuming that intelligent entities themselves possess knowledge of their environments.

[5], p. 2

Nebel riprende ed estende questa posizione:

Knowledge does not seem to be describable in mechanical terms. It seems to be a property we ascribe to an agent when it is *acting*. And this ascription is not just arbitrary or accidental, but rather serves as an important abstraction which permits the explanation of *behavior*.

[10], p. 14, il corsivo è mio

Come preannunciato, in queste definizioni scompare il collegamento con il concetto di dato; inoltre, il collegamento fra K e azione, o comportamento, è chiaramente evidenziato, così come è evidenziato (nella seconda definizione) il fatto che la K è una caratteristica *soggettiva*, e quindi ogni essere umano è padrone di "ascrivere K" (per usare i termini di Nebel) ad agenti differenti: un agente *A* può avere K per *A'* e non averne per *A''*. È quindi importante il ruolo dell'*osservatore*: a seconda di chi osserva, l'ascrizione della conoscenza può avvenire o meno.

Questi due importanti temi (collegamento col comportamento e soggettività) vengono indagati anche da Maturana e Varela, da un punto di vista filosofico-biologico (e quindi non filosofico-IA):

Parliamo di conoscenza ogni volta che osserviamo un comportamento efficace (o adeguato) in un contesto preciso, cioè in un dominio che definiamo con una domanda (esplicita o implicita) che formuliamo come osservatori.

[...] un osservatore può attribuire il valore di atto conoscitivo a qualsiasi interazione realizzata da un organismo e a qualunque comportamento osservato. [...] Si può dire che vivere è conoscere.

[8], p. 147-148

Le considerazioni illustrate dalle ultime tre citazioni hanno portato alla nascita del settore della *vita artificiale* (o AL, da *artificial life*). Tale settore è in rapida crescita (si veda ad esempio [3]) e si basa sull'assunto che per riuscire a riprodurre il comportamento intelligente (e di conseguenza la K) di un essere vivente, è necessario simularne la *vita*, lavorando in modo *bottom-up*, partendo dai processi percettivi, dalla sopravvivenza, dalla riproduzione in un ambiente in continua trasformazione, ecc. Insomma, come sostenuto in [7], l'approccio

classico delle basi di conoscenze non è sufficiente e bisogna "dotare il calcolatore di un corpo". Un'opinione analoga si può in fondo ritrovare anche nelle parole di Searle:

Nessuno pensa che una simulazione al calcolatore della digestione possa realmente digerire qualcosa, ma quando si ha a che fare con i processi cognitivi si è disposti a credere in miracoli del genere perché non ci si rende conto che *la mente è un fenomeno biologico al pari della digestione*. La mente, si suppone, è qualcosa di astratto e formale, non fa parte di quella roba «umida e appiccicaticcia» contenuta nella nostra testa.

[13], p. 21, il corsivo è mio

Si presti attenzione al fatto che, mentre Nebel sostiene che l'ascrizione è soggettiva ma non arbitraria, Maturana e Varela si spingono oltre, sostenendo l'*arbitrarietà* dell'attribuzione dell'«atto conoscitivo». A sostegno di quest'ultima tesi, si può riportare un esempio forse un po' forzato, ma convincente: un termostato ha conoscenza? C'è chi, con una posizione estremistica, sostiene che sì, i termostati "sanno" quando fa troppo caldo, e chi sostiene che un termostato non sa che cosa significhi "fare troppo caldo" (con questo punto di vista concorderebbe sicuramente Searle).² Per inciso, questa considerazione può servire a chiarire un problema che ha dato origine ad accese dispute: l'esperimento mentale della stanza cinese di Searle (si veda [13]). Infatti, se si accetta che la semantica sia una caratteristica, al pari della K, ascrivibile in modo soggettivo, diventa privo di senso chiedersi se la stanza "possieda semantica" o meno, in quanto persone differenti possono attribuire o non attribuire semantica alla stanza cinese basandosi su considerazioni personali. Questo non è più vero se si ha a disposizione una definizione non ambigua (anche se informale) di semantica, su cui tutti concordino; ma abbiamo una tale definizione?

Per concludere questa carrellata di citazioni e contribuire mio malgrado a rinforzare l'opinione di coloro che, forse pessimisticamente, dubitano di poter giungere ad una definizione formale di K, si osservi che uno stesso osservatore può attribuire conoscenze differenti a due agenti differenti, collocati nello stesso ambiente ed esibenti lo stesso comportamento. Ad esempio, è diffusa l'opinione che un programma per gli scacchi in realtà non giochi a scacchi, ma si limiti a simulare (e non a riprodurre) i ragionamenti di uno scacchista umano. Tale opinione è espressa, in modo più generale, da Searle:

[...] è importante rendersi conto che la simulazione non coincide con la riproduzione e l'importanza di questo fatto è la stessa tanto per il pensare di aritmetica quanto per sentire l'angoscia. Non è che il calcolatore arrivi solo fino alla metà del campo invece di arrivare fino all'area di rigore. Il calcolatore non parte neppure: non gioca a questo gioco.

[13], p. 21

3 Conoscenza: un tentativo di definizione

Combinando le ultime definizioni del paragrafo precedente si ha un quadro abbastanza realistico, ma tale quadro può a mio avviso essere reso più nitido e completo. In questo paragrafo effettuo un tentativo in tale direzione, proponendo quello che in seguito chiamerò "definizione di K", ma che in realtà è unicamente un elenco di caratteristiche che la K deve secondo me possedere. Ecco l'elenco:

- la K è una caratteristica che un agente A_1 può ascrivere, attribuire ad un altro agente A_2 ;
- l'attribuzione di K è un'astrazione che permette ad A_1 di spiegare il comportamento di A_2 .
- gli scopi di A_1 giocano un ruolo importante: A_1 vuole comprendere, spiegarsi il comportamento di A_2 e quindi gli attribuisce K;
- tale attribuzione non è forse completamente arbitraria, ma è senz'altro soggettiva;
- tale attribuzione non è un'operazione atomica, in quanto può essere scomposta in 2 operazioni più elementari. Dapprima l'osservatore A_1 si immedesima nell'agente osservato A_2 e immagina di eseguire le stesse azioni (operazione di *immedesimazione*); poi A_1 effettua un'operazione di *introspezione* per spiegare a sé stesso il comportamento che ha osservato nell'altro agente A_2 e che immagina di riprodurre. In parole povere, è come se A_1 si dicesse: «Se io fossi in A_2 , e agissi così, le mie azioni si potrebbero spiegare assumendo che io abbia una certa K»;
- anche l'operazione di immedesimazione può essere ulteriormente analizzata: essa è infatti un caso particolare della funzione di *creazione di un mondo controfattuale*. Nel caso della funzione di immedesimazione, il mondo controfattuale creato è quello in cui l'osservatore A_1 immagina di essere l'agente A_2 ;
- la K va *appresa (learned)*: gli unici esseri che unanimemente si ritiene "abbiano K" sono gli esseri umani e la K è una proprietà dinamica, in continua evoluzione. È comunque importante osservare che l'apprendimento non inizia da una "tabula rasa", ma da una base innata cerebrale, ottenuta per evoluzione della specie.

² Un aspetto che non ho considerato qui e che forse andrebbe tenuto in considerazione è quello della *complessità*: probabilmente si è disposti ad attribuire K solo ad enti che raggiungano un livello minimo di complessità.

Prima di indagare le conseguenze di questa definizione di K sull'IA, riassumo la situazione in modo più formale.

4 Un approccio più formale

La situazione è schematizzata in figura 1: l'agente A_s ("s" deriva da "see") osserva l'agente A_a ("a" da "act"). A_s individua anche un ambiente in cui A_a agisce: tale ambiente è denotato $E_{A_s}(A_a)$ ("E" dall'inglese "environment"), da leggersi "l'ambiente di A_a secondo A_s ". A_s

L'agente A_s osserva le azioni di A_a ; tali azioni costituiscono il *comportamento* di A_a osservato da A_s ; il comportamento è indicato con $B_{A_s}(A_a)$ ("B" da "behaviour"). Al fine di spiegare il comportamento di A_a , A_s gli attribuisce K, denotabile con $K_{A_s}(A_a)$ (la K attribuita da A_s ad A_a).

Si potrebbe supporre che $K_{A_s}(A_a)$ dipenda unicamente dall'ambiente e dal comportamento, ossia, in simboli (indicando con *attr* la funzione di attribuzione di K):

$$K_{A_s}(A_a) = attr(E_{A_s}(A_a), B_{A_s}(A_a)).$$

Ma questo non è vero, come illustrato dagli esempi del programma di scacchi e del termostato presentati alla fine del paragrafo 2.5. Infatti (caso del programma che gioca a scacchi), se si considera un altro agente A_a' , pur essendo che $B_{A_s}(A_a) = B_{A_s}(A_a')$ e $E_{A_s}(A_a) = E_{A_s}(A_a')$, si può avere che

$$K_{A_s}(A_a) \neq K_{A_s}(A_a'),$$

a quindi la $K_{A_s}(A_a)$ dipende anche dall'agente A_a . Analogamente (caso del termostato), se si considerano due osservatori A_s e A_s' , pur avendo $B_{A_s}(A_a) = B_{A_s'}(A_a)$ e $E_{A_s}(A_a) = E_{A_s'}(A_a)$, si può avere che

$$K_{A_s}(A_a) \neq K_{A_s'}(A_a),$$

e quindi la $K_{A_s}(A_a)$ dipende anche dall'osservatore. Riassumendo, la $K_{A_s}(A_a)$ è quindi funzione di agente ed osservatore:

$$K_{A_s}(A_a) = attr(A_a, A_s, E_{A_s}(A_a), B_{A_s}(A_a)).$$

Analizziamo più in dettaglio la funzione *attr*; ho già illustrato nel paragrafo precedente come essa sia la composizione di due funzioni più elementari: le funzioni di introspezione (che indico con *intr*) e di immedesimazione (che indico con *imm*); quest'ultima è inoltre un caso particolare della funzione di creazione di un mondo controfattuale (che indico con *cf*). Quindi si ha che:

$$attr = intr \circ imm = intr \circ cf.$$

La formalizzazione parziale presentata finora lascia molte domande senza risposta: che cosa c'è di diverso in due osservatori che osservano un agente che effettua alcune azioni in un ambiente e che fa sì che le conoscenze ascritte possano essere differenti? Che cosa c'è di diverso in due agenti che effettuano le stesse azioni nello stesso ambiente che fa sì che le conoscenze loro ascritte da un osservatore possano essere differenti (unicamente i rispettivi scopi o altro)? Che cosa vede l'osservatore nell'agente che non possa essere codificato nell'ambiente e

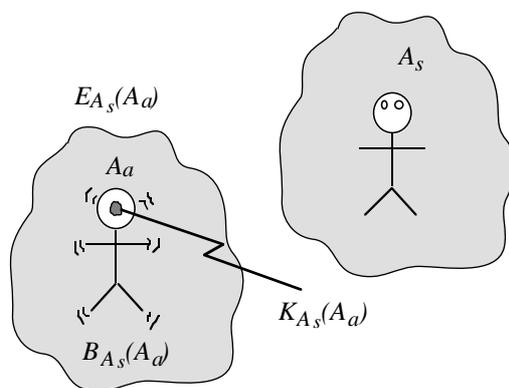


Fig. 1. Schematizzazione dell'operazione di attribuzione di K da parte di A_s ad A_a .

nel comportamento?

Per tentare di rispondere a queste domande si può introdurre il concetto di *modello*: A_s si costruisce un modello di A_a , modello che indico con $M_{A_s}(A_a)$. Tale modello è senz'altro soggettivo (se non arbitrario), quantomeno per la dipendenza da precedenti esperienze di A_s , e potrebbe essere la causa della soggettività (o dell'arbitrarietà) dell'attribuzione di K. La non univocità dell'attribuzione di K potrebbe però essere determinata anche da altri fattori, oltre al modello; in simboli, si avrebbe che

$$M_{A_s}(A_a) = M_{A_s'}(A_a)$$

e

$$K_{A_s}(A_a) \neq K_{A_s'}(A_a).$$

Purtroppo, poter stabilire se questo può effettivamente avvenire è senz'altro difficoltoso, se non impossibile: per farlo, bisognerebbe "entrare nelle teste" di A_s e A_s' , per osservare e confrontare $M_{A_s}(A_a)$ e $M_{A_s'}(A_a)$, e nessuno, almeno per ora, è in grado di effettuare tale operazione.

La formalizzazione dello scenario proposta qui non è completa, e necessiterebbe senz'altro di ulteriore lavoro. Comunque, la strada è tracciata e quindi in questa sede non proseguo su tale linea; accenno invece ai legami esistenti fra quanto detto e il settore dell'IA.

5 L'intelligenza artificiale

La definizione di K del paragrafo 3 può sembrare corretta, ma solleva senz'altro dubbi sulla sua utilità pratica. Comunque, l'utilità di tale definizione si manifesta se la si adoperava per comprendere il lavoro dell'IA, o della KR. In questo paragrafo, dapprima riassumo brevemente il procedimento seguito per la costruzione di una base di conoscenze (paragrafo 5.1) e poi, nel paragrafo 5.2, illustro i punti deboli di tale procedimento alla luce della definizione di K presentata nel paragrafo 3.

5.1 La costruzione di una base di conoscenze

La KR ha l'obiettivo di rappresentare le conoscenze che noi umani utilizziamo per svolgere determinati compiti in modo che esse siano utilizzabili da un calcolatore per riprodurre (o, se si è dell'opinione di Searle illustrata dall'ultima citazione del paragrafo 2.5, simulare) il nostro comportamento. Analizziamo in dettaglio i passi che vengono compiuti in quest'opera di rappresentazione:

- 1) il comportamento di un agente A_a viene osservato. Usando la terminologia del paragrafo 4, si può dire che vi è un agente A_a , che agisce in un suo ambiente, un altro agente A_s che individua l'ambiente $E_{A_s}(A_a)$ e osserva il comportamento di A_a , $B_{A_s}(A_a)$;
- 2) lo scopo di A_s è di spiegare il comportamento di A_a ; a tal fine egli attribuisce ad A_a le $K_{A_s}(A_a)$, ossia le conoscenze sufficienti a spiegare il suo comportamento. Per poter fare ciò, A_s effettua dapprima un'operazione di immedesimazione, ossia crea un mondo controfattuale (il mondo ricavabile da "Se A_s fosse A_a , ..."), e poi un'operazione di introspezione.

5.2 I punti deboli dell'intelligenza artificiale

Nel procedimento descritto nel paragrafo 5.1 si possono individuare alcuni punti deboli:

- l'ambiente $E_{A_s}(A_a)$ è funzione dell'osservatore A_s ;
- il comportamento $B_{A_s}(A_a)$ è funzione dell'osservatore A_s ;
- l'esito dell'operazione di attribuzione della K non è unico. Infatti, tale operazione è di tipo induttivo (indurre una regola generale dall'osservazione di fatti specifici) e, per il principio della povertà dello stimolo, l'esito di tale attribuzione non è unico, in quanto uno stimolo di cui si voglia catturare la regolarità è completabile in infiniti modi diversi;
- l'operazione di costruzione di un mondo controfattuale (componente, come già ripetuto più volte, dell'operazione di immedesimazione) è per sua natura problematica, in quanto non si riescono a definire tutte le condizioni che devono variare rispetto al mondo reale (ciò che si denomina *ceteris paribus* – si veda [9]);
- l'operazione di introspezione è ovviamente soggettiva (ognuno la effettua "a modo suo"), ma è anche problematica in quanto non si può essere certi che ciò che crediamo di "vedere" dentro di noi sia corretto;
- sono completamente assenti ogni forma di apprendimento e di evoluzione della K: è il programmatore (o ingegnere della conoscenza) che fornisce al calcolatore le K "belle e pronte".

Visti e considerati tutti questi problemi, si può senz'altro rilevare quanto l'obiettivo della KR sia ambizioso e

quanto la metodologia adottata sia criticabile; ciò spiega forse perché la KR e, di conseguenza, l'IA siano ben lontani dall'aver raggiunto una condizione stabile e soddisfacente.

6 Conclusioni e lavoro futuro

Gli interrogativi sollevati in queste note (si vedano la fine dei paragrafi 2.4 e 4) sono più numerosi delle risposte fornite; a tali interrogativi si può affiancare una riflessione, che colloca l'indagine sulla K in un ambito più generale. Vi sono vari concetti che noi esseri umani usiamo comunemente, ma che sembrano non definibili formalmente, quali la vita, le intenzioni o la volontà. Io credo che uno studio integrato di questi concetti possa fornire più risposte rispetto ad un approccio unidisciplinare simile a quello esposto qui.

Inoltre, le critiche mosse alla metodologia dell'IA nel paragrafo precedente non sono di tipo propositivo, forse a causa del fatto che l'approccio seguito nel settore della KR è il migliore disponibile, almeno per ora. Comunque, questo lavoro illustra quanto la comprensione del "fenomeno conoscenza" sia limitata ed approssimativa e gli interrogativi proposti, unitamente al completamento della formalizzazione dello scenario delineata nel paragrafo 4, possono essere una utile agenda per migliorare tale comprensione.

Ringraziamenti

Desidero ringraziare il prof. Giuseppe O. Longo, il dott. Paolo Giangrandi e Nello Cristianini, che hanno letto versioni preliminari di questo lavoro e mi hanno fornito preziosi suggerimenti.

Riferimenti bibliografici

- [1] Brachman, R. J. e Levesque, H. J., 1985, *Readings in Knowledge Representation*, Morgan Kaufmann, Los Altos, California.
- [2] Charniak, E. e Mc Dermott, D., 1988, *Introduzione all'intelligenza artificiale*, Masson, Addison-Wesley.
- [3] Fondazione Carlo Erba, in preparazione. *The Future of Science has begun – Approaches to Artificial Life and Artificial Intelligence*, Atti del convegno tenutosi a Milano, il 15-16 Dicembre 1994, organizzato dalla Fondazione Carlo Erba.
- [4] Fum, D., 1994. *Intelligenza artificiale*, il Mulino.
- [5] Genesereth, M. R. and Nilsson, N. J., 1987. *Logical Foundations of Artificial Intelligence*, Morgan Kaufmann, Los Altos, CA.
- [6] Holsheimer, M. e Siebes, A., 1994. *Data Mining: the search for knowledge in databases*, Report CS-R9406 1994, CWI, P. O. Box 94079 GB Amsterdam, The Netherlands.
- [7] Longo, G. O., 1994. *Dal Golem a Gödel e ritorno*, inedito.
- [8] Maturana, H. R., and Varela, F. J., 1987. *L'albero della conoscenza*, Garzanti. Titolo originale: *El árbol del conocimiento*, 1984.
- [9] Mizzaro, S., 1992. *Elaborazione del linguaggio naturale: analisi e inferenze a livello pragmatico*, Tesi di laurea, Università degli Studi di Udine.
- [10] Nebel, B., 1990. Reasoning and Revision in Hybrid Representation Systems, *Lecture Notes in Artificial Intelligence*, n. 422, Springer-Verlag.
- [11] Newell, A., 1982. The knowledge level. In *Artificial Intelligence*, 18(1): 87-127.
- [12] Rich, E., 1983. *Intelligenza Artificiale*, McGraw-Hill, Milano.
- [13] Searle, J. R., 1990. La mente è un programma? In *Le Scienze*, 259, marzo 1990: 16-21.
- [14] Smith, B. C., 1982. *Reflection and Semantics in a Procedural Language*, Tesi di dottorato e Technical Report MIT/LCS/TR-272, M. I. T., Cambridge, MA. Il prologo di tale tesi è ripubblicato in [1], 32-40.
- [15] Sowa, J. F., 1984. *Conceptual structures - Information Processing in Mind and Machine*, Addison-Wesley, Reading, MA.
- [16] Wiederhold, G., 1986. Knowledge versus data. In M. L. Brodie and J. Mylopoulos editors, *On knowledge Base Management Systems: 77-82*, Springer-Verlag, Berlin, Germany.
- [17] Winston, P. H., 1992. *Artificial Intelligence*, Addison-Wesley.
- [18] Wong, H. K. T., and Mylopoulos, J., 1977. Two views of data semantics: data models in artificial intelligence and database management, *INFOR*, 15(3): 344-383.