

Agenti artificiali e agenti intelligenti: paradigmi, applicazioni e prospettive

Anna Goy e Ilaria Torre, Università di Torino

Artificial agents and intelligent agents: paradigms, applications and future trends

English abstract: The concept of *agency* in computer science is related to computational entities, called *artificial agents*, which are able to interact with each other or with human beings. Intelligent agents are specific agents that perform tasks and activities requiring reasoning, thus showing intelligent behaviours. The objective of this paper is providing an overview of this research field, considering its historical development, the proposed approaches and the main applications. The paper discusses a possible definition of agency in the computer science research, presenting the two main approaches to the development of agents in Artificial Intelligence, i.e., goal-driven deliberative agents, and reactive agents. Within these two main research threads, various architectures are discussed, including hybrid ones. Multi-agent systems are also considered, with their emphasis on communication among interacting agents. The paper ends up with a brief presentation of some possible present and future perspectives, with particular attention to the possibilities offered by the Web.

Key-words: agent, artificial, intelligent, multi-agent systems, knowledge representation, artificial intelligence

1. Introduzione

L'articolo nasce da una presentazione tenuta nell'ambito di un seminario interdisciplinare, organizzato alla fine del 2008 dal Centro Interdipartimentale di Ricerca sulla Comunicazione dell'Università di Torino¹, con l'obiettivo di discutere, da punti di vista diversi, il concetto di *agency*. L'obiettivo di quell'intervento era offrire una panoramica introduttiva delle applicazioni che gli scienziati dell'informazione chiamano *agenti intelligenti*. Il presente articolo eredita quell'impostazione e si rivolge quindi prevalentemente a coloro che, non essendo ricercatori e studiosi in ambito informatico, vogliono conoscere il punto di vista delle discipline informatiche su agenti artificiali e agenti intelligenti, contribuendo in tal modo alla costruzione di una visione multidisciplinare del concetto di *agency*.

¹ <http://www.circe.unito.it/>

Con una certa approssimazione si può dire che la disciplina che si è occupata e si occupa di agenti intelligenti in ambito informatico è l'*Intelligenza Artificiale* (anche se non è sempre vero che gli “agenti” artificiali sono anche “intelligenti”, né che tutti i sistemi “intelligenti” sono “agenti”).

L'articolo che presentiamo percorrerà le tappe principali della storia degli agenti intelligenti, cercando di mettere in luce le caratteristiche che, di volta in volta, hanno rappresentato gli aspetti salienti e maggiormente dibattuti dai ricercatori del settore. In questa panoramica, si terrà conto degli approcci che si sono via via susseguiti, ponendo però anche la dovuta attenzione alle ricadute tecnologiche e applicative di tali ricerche. A conclusione dell'articolo si cercherà di capire quale direzione la ricerca sugli agenti intelligenti potrà prendere nel prossimo futuro.

Un'ultima considerazione: la bibliografia sugli agenti intelligenti è estremamente vasta. Il presente articolo si ispira, nella proposta degli argomenti, principalmente a due testi base sugli agenti intelligenti, il testo di Michael Wooldridge *An Introduction to Multi Agent Systems*: cfr. Wooldridge (2002) e quello di Russel e Norvig, *Artificial intelligence: a modern approach*: cfr. Russel e Norvig (2003). Il primo è particolarmente interessante per la sua prospettiva storica, il secondo per l'approccio interdisciplinare con il quale affronta la materia. Un'altra fondamentale fonte bibliografica è Wooldridge e Jennings (1995).

2. Breve sguardo da una prospettiva storica

2.1. Tappe principali

La nascita dell'Intelligenza Artificiale come disciplina di ricerca autonoma si può far risalire al 1956, anno in cui John McCarthy organizzò la prima *Dartmouth Summer Research Conference on Artificial Intelligence* e, contestualmente, ebbe inizio anche la ricerca sugli agenti intelligenti.

A partire dagli anni '60, e fino alla metà degli anni '80 circa, l'Intelligenza Artificiale è stata dominata dal paradigma di *ragionamento e pianificazione* (cfr. Par. 3), caratterizzato da una rappresentazione esplicita e simbolica della conoscenza che l'agente ha del mondo. Questa rappresentazione è generalmente basata su

formalismi logici, grazie ai quali è possibile sfruttare tecniche di ragionamento deduttivo.

Gli anni '70 e '80 vedono una forte crescita del settore e investimenti consistenti per lo sviluppo di *sistemi esperti*, ossia sistemi intelligenti specializzati in compiti diversi, grazie alla vasta base di conoscenza di cui sono dotati. Le eccessive aspettative dei primi anni producono tuttavia una fase successiva di delusione e un conseguente rallentamento di interesse verso questo settore. Viene abbandonato il termine "sistema esperto", in favore di un più generico *sistema basato su conoscenza* e l'Intelligenza Artificiale matura un atteggiamento più realistico e metodologicamente più scientifico.

In questo contesto emerge, alla fine degli anni '80, un nuovo approccio, critico nei confronti dell'approccio simbolico. Il primo e principale esponente di questa critica è Rodney Brooks, che può essere considerato il padre dell'approccio *reattivo* all'Intelligenza Artificiale (*reactive o behavioural Artificial Intelligence*); cfr. Brooks (1991a, 1991b)². Secondo quest'approccio, il comportamento intelligente è di tipo reattivo, emerge cioè dall'interazione di comportamenti semplici e non è supportato da una rappresentazione della conoscenza esplicita e simbolica (cfr. Par. 3).

Negli anni '90 si comincia a ragionare sulla possibilità di integrare i due approcci, producendo sistemi *ibridi* che sfruttano sia il ragionamento deduttivo sia il comportamento reattivo.

Parallelamente, tra gli anni '70 e '80, si sviluppa anche un filone di ricerca sui *sistemi multi-agente* (cfr. Par. 4), nei quali l'enfasi è posta sull'interazione, vista prevalentemente come scambio di messaggi tra agenti artificiali. All'interno di questo filone, verso la metà degli anni '90, vengono compiuti notevoli sforzi volti alla standardizzazione dei linguaggi di comunicazione tra agenti (cfr. Par. 4).

Nel 1990 nasce il World Wide Web e nel 1993 il primo browser ad interfaccia grafica, cfr. Berners-Lee, Cailliau, Luotonen, Frystyk Nielsen e Secret (1994), che rende l'utilizzo di Internet via via sempre più diffuso. Si moltiplicano i siti Web, i servizi offerti sulla Rete e si aprono nuove prospettive anche per la ricerca sugli agenti intelligenti. Da segnalare, nel 1994, uno *Special Issue* della rivista *Communications of the ACM* che ha come argomento gli agenti

² Si veda il dibattito con Oren Etzioni relativo alla necessità che l'ambiente reale circostante debba essere fisico (Brooks) o computazionale (Etzioni); cfr. Etzioni (1993).

intelligenti, visti come *assistenti personali*, che operano soprattutto sul Web per assistere l'utente nella navigazione o nell'utilizzo e nella selezione di servizi. Il principale riferimento in quest'area è rappresentato dai lavori sugli *user interface agents* di Pattie Maes: cfr. Maes (1994).

Infine, grazie ai progressi tecnologici di Internet e allo sviluppo del linguaggio di programmazione Java³ nascono gli *agenti mobili*, programmi capaci di spostarsi sulla Rete ed essere eseguiti in remoto; cfr. Lange e Oshima (1999).

2.2. Campi di applicazione

Il concetto di "intelligenza", e quindi di "agente intelligente", è stato oggetto di studio di molte discipline: la filosofia, la semiotica, la psicologia cognitiva, le neuroscienze, l'economia, la sociologia, ecc. hanno proposto teorie e modelli diversi per analizzare sia l'essenza dell'intelligenza, sia la sua manifestazione. In ambito informatico, la tendenza prevalente è quella di considerare intelligente un agente che esibisce un *comportamento* intelligente. La comprensione del linguaggio naturale, la visione, il riconoscimento di pattern, l'apprendimento, ecc. sono esempi di attività che un agente può compiere, tipicamente in domini di conoscenza specifici, mostrando capacità di ragionamento (deduttivo, induttivo o analogico); cfr. Par. 3.

I campi di applicazione sono moltissimi; si veda, per esempio, Jennings e Wooldridge (1998). Senza pretesa di esaustività, ne citiamo alcuni. La diagnosi medica è stata uno dei primi campi, e di maggior successo, in cui gli agenti intelligenti trovarono applicazione. MYCIN era, per esempio, un sistema per diagnosticare le infezioni del sangue: cfr. Shortliffe (1976). Dotato di una base di conoscenza molto vasta "poteva offrire prestazioni pari a quelle di molti esperti"; Russel e Norvig (2003), p. 33. Altri campi di applicazione sono quelli di *business process management* (si veda per esempio il progetto ADEPT di British Telecom: cfr. Jennings, Norman e Faratin 1998), di gestione del traffico aereo (il progetto OASIS ne è un esempio, con sistemi multi-agenti che supportano i controllori di volo: cfr. Ljungberg e Lucas 1992), dei giochi (Deep Blue è stato il primo sistema in grado di sconfiggere il campione Garry Kasparov nel gioco degli scacchi; cfr. Goodman e Keene 1997), ecc.

³ <http://java.sun.com/>

Con lo sviluppo del Web nascono nuovi possibili ambiti di applicazione degli agenti intelligenti. Nel campo dell'*Information Retrieval*⁴, per esempio, gli agenti vengono usati per recuperare informazioni da sorgenti eterogenee, filtrarle in modo personalizzato e organizzarle in una visione coerente (MAXIM, per esempio, è un *email agent* che impara osservando il comportamento dell'utente: cfr. Maes, 1994). Nel commercio elettronico è frequente l'uso di agenti, spesso chiamati (*ro*)bots, che supportano vari aspetti della transazione, quali la ricerca del prodotto (con comparazioni multi-attributo di prodotti e produttori), la partecipazione ad aste, la negoziazione del contratto e la personalizzazione: cfr. Goy, Ardissono e Petrone (2007).

3. Che cos'è un agente intelligente?

3.1. Un tentativo di definizione

Non esiste una definizione accettata e condivisa di cosa sia un agente intelligente: il dibattito è aperto e vede intersecarsi, e talvolta anche contrapporsi, definizioni diverse; si veda, per esempio, Franklin e Graesser (1997). Una definizione che potrebbe trovare una certa convergenza è la seguente.

Un agente è un sistema computazionale, costituito da un programma software ed eventualmente da un supporto hardware, che:

- interagisce con l'ambiente circostante ed è *reattivo* agli stimoli di tale ambiente;
- è capace di prendere decisioni, e di conseguenza di agire, in modo *autonomo*, con il fine di raggiungere un obiettivo, chiamato generalmente *goal* (che può essere predefinito o negoziato); pertanto può essere definito *proattivo*;
- è in grado di *comunicare* (coordinarsi, cooperare, negoziare) con altri agenti (e/o con esseri umani), è quindi capace di *interazione sociale*.

L'ambiente circostante in cui agisce l'agente può essere un ambiente fisico (reale o artificiale, quale un set di laboratorio), un ambiente computazionale (ossia un ambiente software, quale il

⁴ L'*Information Retrieval* è l'attività di recupero di informazioni da vasti corpora testuali non strutturati, come per esempio il Web. E' la principale tecnologia che sta alla base del funzionamento dei motori di ricerca.

sistema operativo del computer o Internet) o un ambiente virtuale (ossia un ambiente fisico costruito in modo virtuale mediante tecnologie hardware e software di *realtà virtuale*). Nel primo caso l'agente sarà un *robot*, nel secondo caso sarà un *agente software* e nel terzo un *agente artificiale*.

Altre proprietà che alcuni ascrivono agli agenti, ma rispetto alle quali non c'è accordo unanime, sono la mobilità e la capacità di apprendimento. Per *mobilità* intendiamo la capacità dell'agente di muoversi nel mondo fisico (se si tratta di un robot), oppure di muoversi sulla Rete (se si tratta di un agente software). Per capacità di *apprendimento* (*machine learning*) intendiamo la capacità di un agente di acquisire nuove conoscenze, o migliorare le proprie prestazioni, sulla base dell'esperienza. A seconda di ciò che si vuole apprendere e del tipo di feedback disponibile, l'apprendimento può assumere forme diverse. Per esempio, si può apprendere l'utilità di un comportamento in base agli effetti che produce, si possono apprendere descrizioni di categorie a partire da una serie di casi, o si può anche imparare a riconoscere pattern o schemi in assenza di informazioni sull'output.

3.2. Approccio deliberativo e agenti basati su obiettivi

Gli agenti basati su approcci di tipo simbolico-deliberativo sono agenti dotati di conoscenza che permette loro di agire e raggiungere i propri obiettivi. L'agente deve possedere conoscenza di tipo sia dichiarativo sia operativo. In termini generali diciamo che la conoscenza *dichiarativa* definisce le entità del mondo (*miniworld*), in cui l'agente agisce (dominio applicativo), specificandone proprietà e relazioni. La conoscenza *operativa*, o *normativa*, è una conoscenza tipicamente definita in termini di regole, che consente all'agente di agire trasformando il mondo in stati successivi, dallo stato iniziale allo stato finale, rappresentato dai goal dell'agente.

Il funzionamento di un agente basato su obiettivi può essere descritto utilizzando il *modello BDI* (*Beliefs, Desires, Intentions*), che si basa su concetti provenienti dall'area di ricerca delle scienze cognitive⁵. Secondo questo modello, un agente possiede: (a) delle

⁵ Il modello BDI è stato implementato con diverse architetture software e trova il suo fondamento nella teoria del *practical reasoning* di Michael Bratman e colleghi;

credenze (*beliefs*), ossia assunzioni relative al proprio stato interno e all'ambiente in un determinato istante; (b) dei desideri (*desires*), cioè degli stati del mondo che vorrebbe che si producessero; (c) delle intenzioni (*intentions*), ossia quei desideri che sono diventati obiettivi da raggiungere, attraverso la selezione di opportune azioni. Le intenzioni rappresentano lo stato deliberativo dell'agente.

Ma come si costruisce un agente intelligente? Come si fa ad "insegnare" ad un agente cosa deve fare? Come si fa a dotarlo di "intenzioni", a "spiegargli" quali sono i suoi obiettivi (*goal*) e a dotarlo di capacità di azione per raggiungerli?

Secondo l'approccio simbolico-deliberativo, basato sul *paradigma di ragionamento e pianificazione* (cfr. Par 2.1) – cfr. Rao e Georgeff (1991), è necessario fornire all'agente una descrizione formale:

- dell'ambiente circostante (in termini di stati), che rappresenta la *conoscenza* che l'agente ha del mondo;
- del goal (uno stato), che rappresenta l'*intenzione* dell'agente, nei termini del modello BDI;
- delle azioni che è possibile effettuare nell'ambiente per raggiungere il goal, corredate, ognuna, di precondizioni ed effetti (entrambi forniti in termini di stati dell'ambiente).

Storicamente, il formalismo più utilizzato per fornire tale rappresentazione è stato la Logica, formalizzata da Boole a metà dell'800. Da questa sono poi stati derivati altri formalismi che hanno tentato di produrre rappresentazioni più compatte o computazionalmente più efficienti, o che hanno infine introdotto componenti procedurali, volti all'azione. A seconda di quali strumenti e metodologie specifiche si usano per fornire la conoscenza al sistema, si possono distinguere due principali tipi di agenti: gli agenti cosiddetti *logici* e gli agenti chiamati *pratici*.

Negli agenti logici, la rappresentazione dell'ambiente, del goal e delle possibili azioni è costituita da un insieme di formule logiche e la manipolazione di tale rappresentazione prende la forma di una deduzione (prova di teoremi).

La costruzione di agenti pratici invece si ispira ai processi cognitivi umani e cerca di realizzare quello che viene chiamato *ragionamento pratico*, diretto all'azione; cfr. Bratman (1987) e Bratman, Israel e Pollack (1988). Il ragionamento pratico si articola in due passi. Il

cfr. Bratman (1987), Bratman, Israel e Pollack (1988). Si veda anche Rao e Georgeff (1991).

primo passo è rappresentato dalla scelta dell'obiettivo, cioè dello stato da raggiungere; scegliere lo stato e decidere di raggiungerlo corrisponde a creare un *commitment*: il goal diventa così un'intenzione. Il secondo passo è rappresentato dal cosiddetto *ragionamento mezzi-fini*, che corrisponde alla *pianificazione*: l'agente stabilisce il piano (mezzo) per raggiungere il goal (fine). Infine il piano viene tradotto in azione, cioè eseguito nell'ambiente circostante. La fase di pianificazione è guidata da un *algoritmo di pianificazione*, che molto spesso è costituito dalla selezione del piano più opportuno all'interno di una "libreria" di piani (sequenze strutturate di azioni, spesso preconfezionate); cfr. Pollack (1992).

L'esempio più famoso di prototipo progettato e realizzato secondo quest'approccio è il *Procedural Reasoning System* (PRS) di Georgeff e Lansky; cfr. Georgeff e Lansky (1987). Un agente basato su PRS è dotato di un insieme di *credenze*, di un *goal* e di una *libreria di piani* preconfezionati, ognuno dei quali è caratterizzato da un *goal* (post-condizione del piano), un *contesto* (pre-condizione del piano), un *body* (la sequenza di azioni che conducono dal contesto al goal). Quando il programma che implementa tale agente viene avviato, il goal viene inserito in cima a una pila (*stack*) di goal. L'agente cerca nella libreria dei piani se c'è qualche piano che gli permetta di raggiungere il goal in cima alla pila: l'insieme dei piani le cui pre-condizioni sono soddisfatte dalle sue credenze sul mondo e le cui post-condizioni coincidono con il goal da raggiungere rappresentano le *opzioni* disponibili. Attraverso un *ragionamento*, l'agente sceglie il piano da eseguire. Nella struttura originaria di PRS, tale ragionamento era realizzato attraverso *meta-piani*, cioè piani che avevano l'obiettivo di selezionare piani. Tale approccio è stato poi sostituito dal calcolo di una *funzione di utilità* che assegna un valore numerico a ciascun piano presente nell'insieme delle opzioni: l'agente sceglierà il piano con il più alto valore di utilità.

3.3. Approccio comportamentista e agenti reattivi

L'approccio simbolico-deliberativo è stato criticato principalmente da due punti di vista. Innanzi tutto è stata sottolineata la necessità di abbandonare i domini applicativi semplificati tipici degli approcci classici (*toy problems*, problemi giocattolo), per passare alla risoluzione di problemi più complessi, immersi nel mondo reale (*real life problems*) e spesso dominati dall'incertezza. La seconda critica, in

parte connessa alla precedente, contesta la scarsa considerazione dell'ambiente circostante sostenendo che il comportamento intelligente è indissolubilmente legato all'interazione (fisica) con esso e che il comportamento intelligente emerge dall'interazione di comportamenti più semplici e non dalla manipolazione di rappresentazioni simboliche. Queste critiche hanno dato vita agli approcci chiamati *comportamentismi, situati o reattivi*.

L'agente reattivo sceglie le azioni sulla base della percezione corrente. A ogni input percettivo corrispondono delle azioni che vengono scatenate da tale input. Queste azioni possono essere definite utilizzando regole (chiamate *situazione-azione, produzioni* o regole *if-then*) oppure funzioni matematiche e probabilistiche che usano conoscenza sub-simbolica. L'esempio più famoso di sistema progettato secondo questi principi è la *Subsumption Architecture* di Brooks; cfr. Brooks (1986). La *Subsumption Architecture* è costituita da un insieme di "comportamenti" strutturati su livelli gerarchici: ogni comportamento è in sostanza una funzione che riceve input percettivi e li mette in corrispondenza con delle azioni. Ogni funzione (comportamento) è specializzata nell'esecuzione di un particolare compito elementare e può essere implementata come un *automa a stati finiti*⁶.

Secondo i sostenitori di quest'approccio, i principali vantaggi dei sistemi reattivi consistono nell'essere semplici, economici, computazionalmente efficienti e robusti. Gli approcci comportamentisti e reattivi presentano tuttavia anche alcuni svantaggi. In particolare, le relazioni tra i comportamenti semplici, le relazioni con l'ambiente e il comportamento del sistema nel suo complesso, non essendo rappresentati esplicitamente, sono difficilmente comprensibili, spiegabili. Per la stessa ragione, la costruzione di questi sistemi richiede una metodologia per prove ed errori, nel tentativo di far emergere il comportamento desiderato, con conseguente difficoltà nel caso di sistemi che affrontano problemi complessi.

Alla luce di questi limiti, la ricerca si è direzionata verso *agenti ibridi*, che cercassero di esibire tanto un comportamento emergente reattivo, quanto una qualche forma di proattività. L'approccio più semplice per costruire sistemi ibridi è avere diversi sotto-sistemi,

⁶ Gli automi a stati finiti sono sistemi che evolvono nel tempo, passando (in modo discreto) da uno stato all'altro in funzione dei valori in ingresso e dello stato precedente; i possibili valori in ingresso e in uscita sono insiemi finiti.

ognuno dei quali basato su una differente tecnologia; tali sotto-sistemi possono poi essere strutturati in *livelli*, dando così origine alle cosiddette architetture a livelli (*layered architectures*). Minimalmente, i livelli saranno due, uno che opera in modo reattivo e uno che opera in modo proattivo; è tuttavia possibile immaginare un numero maggiore di livelli. Il tipo di interazione tra i livelli dà origine a due diverse tipologie: le architetture “orizzontali”, in cui i livelli operano, per così dire, in parallelo e concorrono a generare l’output; e le architetture “verticali”, in cui ogni livello esegue un’elaborazione dell’input e poi passa il controllo al livello successivo, fino al livello che genera l’output. Le architetture orizzontali sono concettualmente più semplici, ma necessitano, generalmente, di un “coordinatore super-partes” che garantisca la coerenza dell’elaborazione effettuata dai vari livelli.

Due esempi di agenti che seguono l’approccio ibrido sono SOAR – cfr. Laird, Newell e Rosenbloom (1987) e THEO – cfr. Mitchell (1990): ogni volta che risolvono un problema attraverso un ragionamento (simbolico), salvano a parte una versione generalizzata della soluzione affinché il componente reattivo possa utilizzarla – cfr. Russel e Norvig (2003).

Un tipo particolare di approccio comportamentista/reattivo è rappresentato dai *modelli connessionisti* (o *reti neurali artificiali*) che, nati alla fine degli anni ’60, hanno però conosciuto un forte sviluppo negli anni ’80. Tali modelli si ispirano al funzionamento delle reti neurali naturali, e quindi al funzionamento del cervello umano. Il connessionismo nasce come “approccio computazionale e simulativo allo studio della mente e del cervello” – cfr. Parisi (1989), p.12 – inquadrandosi quindi nel filone di quelle critiche all’Intelligenza Artificiale classica secondo le quali il substrato fisico che supporta l’intelligenza (il cervello umano) non poteva essere trascurato. Pur partendo da questi presupposti, tuttavia, i sistemi connessionisti effettivamente sviluppati negli anni, sono rimasti implementazioni di modelli computazionali, ispirati forse (almeno in parte) al funzionamento del cervello, ma di natura non diversa rispetto a quelli simbolico-deliberativi e soprattutto a quelli reattivi, visti sopra.

Per capire di cosa stiamo parlando, descriviamo molto brevemente le caratteristiche di base di un modello connessionista. Una rete neurale è formata da un certo numero di *unità* (o *nodi*), connesse tra loro da *connessioni*; ogni connessione ha un *peso* che rappresenta la capacità di quella connessione di trasmettere *attivazione* (pesi positivi) o

inibizione (pesi negativi) da un'unità all'altra, determinando così lo *stato di attivazione* dell'unità "di arrivo". Le unità di una rete neurale sono organizzate in *strati*, in modo tale che le unità appartenenti ad uno stesso strato non sono connesse tra loro, ma sono connesse allo strato precedente (da cui ricevono input) e allo strato successivo (a cui forniscono output); il primo strato contiene le *unità di input* che ricevono input dall'ambiente, mentre l'ultimo strato contiene le *unità di output* che forniscono output all'ambiente; cfr. Parisi (1989), Ripley (1996).

Una delle caratteristiche principali dei sistemi connessionisti è la loro capacità di apprendimento, che consiste nel modificare i pesi delle connessioni che regolano il propagarsi dell'attivazione nella rete. La conoscenza del sistema – in modo analogo a molti sistemi reattivi – è, quindi, implicita nella configurazione strutturale della rete.

Le reti neurali artificiali sono state utilizzate per moltissimi scopi e in generale sono utili nelle situazioni in cui non siano chiare a priori le relazioni tra le variabili che definiscono il problema. Per esempio, sono state utilizzate per il riconoscimento di cifre scritte a mano, nella visione artificiale, nella classificazione di immagini e nella diagnosi medica. Un'altra area in cui i sistemi connessionisti, spesso combinati in approcci ibridi, hanno trovato applicazione è la *robotica*, disciplina che si occupa di agenti in grado di manipolare il mondo fisico e di conseguenza applica molte delle funzionalità viste sopra. A seconda della tipologia e dei compiti del robot,⁷ possono essere richieste abilità quali la visione artificiale, la comprensione del linguaggio naturale e il movimento in spazi fisici.

4. Sistemi multi-agente e comunicazione tra agenti

Un filone di ricerca per certi versi distinto ed indipendente è quello che si occupa dei cosiddetti *sistemi multi-agente*. Con un po' di approssimazione, possiamo dire che un *multi-agent system* (MAS) è un sistema composto da più agenti che interagiscono tra loro e che generalmente comunicano scambiandosi messaggi.

⁷ Esistono tre tipi principali di robot: i manipolatori o bracci mobili, spesso usati in catene di montaggio industriali; i robot mobili, capaci di spostarsi nell'ambiente ed usati, per esempio, come carrelli "intelligenti" in ospedali o centri di stoccaggio; i robot umanoidi, con capacità sensoriali e attuatori che emulano quelli umani.

Un'architettura multi-agente può definire la struttura interna di un sistema, oppure può modellare una società di agenti che svolgono funzioni di alto livello e coprono ruoli diversi. In ogni caso, ogni agente interagisce con l'ambiente circostante in modo autonomo ed è "egoista", ha cioè le proprie "preferenze" relativamente agli stati del mondo. Di conseguenza, il principale problema all'interno di un'architettura multi-agente è quello di come fare a raggiungere un accordo, un obiettivo comune. In altre parole, il problema è come realizzare la *cooperazione* tra gli agenti.

Per raggiungere tale accordo diventa essenziale la *comunicazione* tra gli agenti del sistema. Nella tradizione di ricerca dell'Intelligenza Artificiale, questa comunicazione è stata modellata seguendo la teoria degli atti linguistici (*speech acts theory*) – cfr. (Searle 1969) che concepisce il linguaggio come azione e trova il suo fondamento teorico nella filosofia del linguaggio del XX secolo. L'idea fondamentale è quella secondo cui i piani dei singoli agenti, che modellano le possibili azioni, includano anche azioni comunicative. Tali azioni, dovendo quindi essere rappresentate in modo che un pianificatore possa ragionarci sopra, vengono rappresentate come operatori, corredati di precondizioni ed effetti che impattano, non su stati del mondo, ma su credenze e goal degli interlocutori; cfr. Cohen e Perrault (1979).

Diversi sono i linguaggi formali proposti per la rappresentazione e lo scambio di messaggi tra agenti, fondati, o quanto meno ispirati, alla teoria degli atti linguistici. In particolare, all'inizio degli anni '90 ha visto la luce l'iniziativa denominata *Knowledge Sharing Effort*, sostenuta dal DARPA⁸. Tale iniziativa produsse due importanti linguaggi standard:

- KQML (*Knowledge Query and Manipulation Language*), che definisce il formato dei messaggi; Finin, Fritzson, McKay e McEntire (1994).
- KIF (*Knowledge Interchange Format*), per la rappresentazione del contenuto⁹.

Nel 1995 è nata la *Foundation for Intelligent Physical Agents* (FIPA), a cui si deve la definizione del linguaggio FIPA-ACL (FIPA

⁸ *Defense Advanced Research Projects Agency* è un'agenzia governativa del Dipartimento della Difesa degli Stati Uniti, che ha in carico lo sviluppo di nuove tecnologie ad uso militare.

⁹ <http://logic.stanford.edu/kif/specification.html>

Agent Communication Language), che specifica il formato dei messaggi,¹⁰ superando il linguaggio KQML, nonché lo sviluppo, nel 1997, del programma *Ontolingua Server* per la costruzione di ontologie¹¹ scritte utilizzando il linguaggio KIF.

La ricerca sui sistemi multi-agente si è infine concentrata su tematiche di *negoziazione* (tecniche per far sì che diversi agenti raggiungano un accordo) e *coordinazione* (tecniche per gestire le inter-dipendenze tra le azioni di diversi agenti), generando altre proposte di approcci e protocolli; cfr. Rosenschein e Zlotkin (1994), Kraus (1997); vonMartial (1992).

5. Tendenze attuali e future

Abbiamo visto quali sono stati i problemi affrontati, gli approcci e le soluzioni proposte nell'ambito della ricerca sugli agenti intelligenti, cercando di delineare l'evoluzione storica di tale ricerca. Ma a che punto siamo oggi? Quali sono le tendenze attuali e le prospettive future? Citiamo, in questa sezione, alcuni filoni di ricerca che si intersecano con lo studio degli agenti intelligenti e che ne indicano, probabilmente, possibili direzioni di sviluppo, soprattutto in ambito Web.

Un primo filone di ricerca è quello legato al Semantic Web Berners-Lee, Hendler e Lassila (2001), finalizzato a costruire un Web in cui agenti artificiali abbiano la capacità di comprendere, e non solo analizzare sintatticamente o in modo statistico, i contenuti e ragionare su questi. Componente base di questo paradigma è la rappresentazione della conoscenza mediante *ontologie* (cfr. nota 7), grazie alle quali i contenuti sono descritti semanticamente utilizzando specifici linguaggi formali.

Internet e il Web hanno anche reso possibile una partecipazione sempre maggiore degli utenti nella creazione di contenuti. Sul Web la conoscenza è sempre più creata in modo collaborativo: questo aspetto ha dato vita al fenomeno del cosiddetto Social Web e alla creazione di

¹⁰ <http://www.fipa.org/repository/aclspecs.html>

¹¹ In informatica, e nel campo dell'Intelligenza Artificiale, per *ontologia* si intende una descrizione formale dei concetti e delle relazioni tra questi, comprensibile da un agente o una comunità di agenti.

*folksonomie*¹². La combinazione di questo fenomeno con il paradigma sopra descritto dovrebbe condurre a una nuova generazione di Web intelligente e sociale, al quale inizia ad essere attribuito il nome di Web 3.0¹³

Citiamo infine un filone che deriva dall'evoluzione degli *Interfaces Agents* di Maes (cfr. Par 2.2) e riguarda lo sviluppo di veri e propri *assistenti personali* per l'uso del computer e della Rete. In questi agenti è diventato di fondamentale importanza l'apprendimento delle preferenze della persona che assistono, a partire dall'osservazione del suo comportamento: questo è possibile grazie all'integrazione, negli assistenti personali, di una componente di *apprendimento automatico*. E' importante sottolineare come la realizzazione di questi agenti, che agiscono prevalentemente (ma non solo) sul Web, attinga competenze anche dall'area di ricerca dell'Interazione Uomo-Macchina.

Concludiamo proponendo alcune riflessioni che hanno l'obiettivo di aprire la discussione su cosa può essere chiamato "agente" ("intelligente") oggi. Non hanno la pretesa di dire una parola conclusiva sull'argomento, ma, al contrario, hanno l'obiettivo di stimolare ulteriori riflessioni e discussioni, in particolar modo in una prospettiva interdisciplinare.

Soprattutto in passato, i termini "agente" e "intelligente" sono stati utilizzati, sia congiunti sia separati, in un'accezione molto simile a quella con cui li si usa nei linguaggi naturali: erano concepiti come entità che non solo esibivano, ma possedevano una qualche forma di intelligenza. Nel tempo questi termini si sono specializzati per assumere significati più precisi, quanto meno all'interno della comunità di ricerca. Se la principale caratteristica distintiva di un agente sembra essere la sua capacità di comunicare, oggi si tende a distinguere:

- agenti con capacità di comunicare con altri agenti nelle cosiddette architetture multi-agente;
- agenti con capacità di comunicare con l'utente mediante interfacce intelligenti.

Nel primo caso, quindi, l'enfasi è sui meccanismi di comunicazione automatici tra agenti, mentre nel secondo caso è sulla

¹² Per *folksonomia* si intende una classificazione di informazioni prodotta dagli utenti del Web, attraverso l'attribuzione di parole chiave (*tag*) ai contenuti.

¹³ Si veda per esempio <http://www.cs.pitt.edu/~rosta/EICK/>

capacità di svolgere attività che richiedono intelligenza al fine di comunicare con, e quindi supportare, l'utente.

Sono poi naturalmente possibili approcci ibridi, che richiedono tuttavia una complessità tale da renderli ancora un obiettivo aperto in questo campo di ricerca: società multi-agente, definite in termini di organizzazioni in cui gli agenti coprono ruoli diversi, interagiscono reciprocamente sia tra di loro, sia con agenti umani per svolgere compiti complessi devono ancora trovare i modelli e i linguaggi adeguati. Concludiamo con una citazione dalla presentazione del decimo workshop internazionale *Engineering Societies in the Agents' World*¹⁴ del 2009, che ci pare delineare in modo efficace questa prospettiva, evidenziando come lo sviluppo di questi sistemi complessi debba necessariamente avvalersi di modelli e teorie provenienti da altri settori disciplinari, legati, ciascuno con peculiarità proprie, al concetto di *agency*:

An urgent need exists for novel approaches to software modelling and software engineering that enable the successful deployment of software systems made up of a massive number of autonomous components, and that allow to control and predict their behaviour. It is very likely that such innovations will exploit lessons from a variety of different scientific disciplines, such as sociology, economics, organisation \science, modern thermodynamics, and biology.

Riferimenti bibliografici

- Berners-Lee T., R. Cailliau, A. Luotonen, H. Frystyk Nielsen e A. Secret (1994) *The World-Wide Web*, "Communications of the ACM", 37(8), 76-82
- Berners-Lee T., J. Hendler e O. Lassila (2001) *The Semantic Web*, "Scientific American Magazine", 284(5), 34-43
- Bratman M. E. (1987) *Intention, Plan and Practical Reason*, Harvard University Press
- Bratman M. E., D. J. Israel e M. E. Pollack (1988) *Plans and resource-bounded practical reasoning*, "Computational Intelligence", 4: 349-355
- Brooks R. A. (1986) *A robust layered control system for a mobile robot*, "IEEE Journal of Robotics and Automation", 2: 14-23

¹⁴ <http://www.cs.uu.nl/esaw2009/index.php>

- Brooks R.A. (1991a) "Intelligence without reason", in *Proc. of the International Joint Conference in Artificial Intelligence (IJCAI-91)*: 569-595
- Brooks R.A. (1991b) *Intelligence without representation*, "Artificial Intelligence", 47: 139-159
- Cohen P. R. e C. R. Perrault (1979) *Elements of a plan-based theory of speech acts*, *Cognitive Sciences*, 3(3):177-212
- Etzioni O. (1993) *Intelligence without robots*, "AI Magazine", 14(4), 7-13
- Finin T., R. Fritzson, D. McKay e R. McEntire (1994) "KQML as an agent communication language", in *Proceedings of CIKM-94*, ACM Press, 456-463
- Franklin S. e A. Graesser (1997) "Is it an agent, or just a program?", in J.P. Müller, M. Wooldridge e N.R. Jennings (a cura di), *Intelligent Agents III*, LNAI 1193, Springer, 21-36
- Georgeff M.P. e A.L. Lansky (1987) "Reactive reasoning and planning", in *Proceedings of AAI-87*, AAAI Press, 677--682
- Goodman D e R. Keene (1997) *Man versus Machine: Kasparov versus Deep Blue*. H3 Publications, Cambridge, MA
- Goy A., L. Ardissono e G. Petrone (2007) "Personalization in e-commerce applications", in P. Brusilovsky, A. Kobsa e W. Nejdl (a cura di), *The Adaptive Web*, LNCS 4321, Springer, 485-520
- Jennings N. R., T. J. Norman e P. Faratin (1998) *ADEPT: an agent-based approach to business process management*. "ACM SIGMOD Record", 27(4): 32-39.
- Jennings N.R. e M. Wooldridge (a cura di) (1998) *Agent Technology: Foundations, Applications and Markets*, Springer
- Kraus S. (1997) *Negotiation and cooperation in multi-agent environments*, "Artificial Intelligence", 94(1-2): 9-98.
- Laird J. E., A. Newell e P.S. Rosenbloom (1987) *SOAR: an architecture for general intelligence*, "Artificial Intelligence", 33(1):1-64
- Lange D.B. e M. Oshima (1999) *Seven Good Reasons for Mobile Agents*, "Communications of ACM", 42(3): 88-89
- Ljungberg M. e A. Lucas (1992) "The OASIS air-traffic management system", in *Proceedings of PRICAI '92*, Seoul, Korea
- Maes P. (1994) *Agents that reduce work and information overload*, "Communications of ACM", 37(7): 31-40
- Mitchell T.M (1990) "Becoming increasingly reactive (mobile robots)" *In Proceedings of the Eight National Conference on*

- Artificial Intelligence (AAAI, 90)*, Vol 2, 1051-1058, Boston, MIT Press
- Parisi D. (1989) *Intervista sulle reti neurali*, Il Mulino
- Pollack M.E. (1992) *The use of plans*, "Artificial Intelligence", 57(1): 43-68
- Rao A.S. e M.P. Georgeff (1991) "Modeling rational agents within a BDI-architecture", in *Proceedings of KR-91*, Morgan Kaufmann, 473-484
- Ripley B. D. (1996) *Pattern Recognition and Neural Networking*, Cambridge University Press, Cambridge, UK
- Rosenschein J.S. e G. Zlotkin (1994) *Rules of encounter: designing conventions for automated negotiation among computers*, MIT Press
- Russel S. e P. Norvig (2003) *Artificial intelligence: a modern approach*, Vol 1 e 2, Prentice Hall (trad. it. *Intelligenza Artificiale: un approccio moderno*, Pearson, Milano 2005)
- vonMartial F. (1992) *Coordinating plans of autonomous agents*, LNAI 610, Springer
- Searle (1969) *Speech Acts: an Essay in the Philosophy of Language*, Cambridge University Press, Cambridge, UK
- Shortliffe H. (1976) *Computer-Based Medical Consultations: MYCIN*, Artificial Intelligence Series, Elsevier
- Wooldridge M. (2002) *An Introduction to Multi Agent Systems*, John Wiley & Sons
- Wooldridge M. e N.R. Jennings (1995) *Intelligent agents: Theory and practice*, "The Knowledge Engineering Review", 10(2): 115-152