Modelli cognitivi nella collaborazione uomo-robot: un approccio basato su delega e adozione

Filippo Cantucci, Rino Falcone, Cristiano Castelfranchi
Istituto di scienze e tecnologie della cognizione (ISTC-CNR), Roma
filippo.cantucci@istc.cnr.it, rino.falcone@istc.cnr.it, cristiano.castelfranchi@istc.cnr.it

Abstract

In questo lavoro presentiamo la prima parte di un progetto di robotica cognitiva che mira allo sviluppo di un modello computazionale in grado di favorire una collaborazione effettiva e basata su fiducia tra uomo e robot. Il modello è progettato per rendere un robot capace di adattare il proprio livello di autonomia collaborativa, sulla base di vari elementi, espandendo o restringendo il compito delegato da un utente. Partendo da concetti di BDI modeling agent, teoria della mente e teoria di delega e adozione, è possibile realizzare un sistema decisionale adatto a contesti reali. Il modello è stato implementato sfruttando il framework *JaCaMo*, che fornisce supporto alla implementazione di sistemi multi-agente, integrando diverse dimensioni di multi-agent programming. Questo modello computazionale è stato testato su un robot reale, il robot umanoide *Nao*, molto utilizzato in esperimenti di human-robot interaction.

1 Introduzione

Nella vita quotidiana gli uomini cooperano tra di loro per scambiarsi conoscenza, condividere obiettivi, e lo fanno seguendo consuetudini sociali che regolano il flusso di collaborazione. Una delle sfide principali in ambito AI, è quella di sviluppare sistemi autonomi in grado di cooperare in modo efficiente e concreto con umani. Ad esempio, i robots stanno entrando ogni giorno di più in ambienti popolati da persone, inclusi ospedali, uffici, classi, e in tali contesti devono coesistere con un ampio spettro di utenti con limitate competenze tecniche. E' sempre più necessario sviluppare sistemi autonomi di cui potersi fidare, capaci di cooperare in modo intelligente, ossia adattando il loro supporto di volta in volta, così come succede naturalmente tra gli umani, interpretando la collaborazione in una visione di teoria della mente e di situazioni contestuali. Negli anni sono state proposte molte architetture cognitive [Kajdocsi e Pozna, 2014] , ognuna con l'obiettivo di simulare aspetti comportamentali e cognitivi tipici dell'uomo, a diversi livelli di cognizione, dalla percezione, all'apprendimento, al reasoning, etc. Tuttavia, oltre alle abilità di interpretare autonomamente il contesto in cui si svolge l'interazione, reagendo a cambiamenti nell'ambiente, e prendere decisioni proattive, i robots dovrebbero mostrare la capacità di seguire le stesse consuetidini sociali che vengono seguite dagli umani quando interagiscono tra di loro, e sfruttare le proprie capacità, nel meccanismo implicato dal concetto stesso di cooperazione. Come sostenuto formalmente in [Castelfranchi e Falcone, 1998], la cooperazione si traduce tipicamente nella delega di un compito τ, da parte di un agente A verso un agente B. L'agente A, il *client*, delega il compito attraverso una specifica richiesta, in termini di offerta, proposta, annuncio etc. all'agente B, il *contractor*, che si impegna ad adottare il compito delegato modulando autonomamente il proprio livello di adozione sulla base di differenti gradi di aiuto, che dipendono dalla contesto. La nozione di autonomia in un agente artificiale, un robot ad esempio, dovrebbe essere fondata su diversi livelli di delega e adozione. Concentrandoci sui livelli di adozione, un robot potrebbe autonomamente modulare il proprio livello di adozione secondo i diversi livelli di help:

- **Sub help**: Il contractor assolve solo parte del compito (parte dello scopo) che il client ha delegato.
- **Literal help**: indica l'assolvimento preciso del compito che il client ha delegato al contractor. Non ci sono modifiche, né sulle strategie né sulle azioni che il client ha deliberato per svolgere il compito che porti al raggiungimento anche parziale dello scopo proposto.
- **Over help**: Il contractor svolge più di quanto richiesto dal client (un suo sovrascopo) e soddisfa in tal modo anche lo scopo delegato (il cui piano realizzativo non viene modificato).
- **Critical help**: Il contractor soddisfa il compito richiesto dal client ma alterando l'intero piano delegato.
- **Critical-Over help**:Il contractor decide autonomamente di evadere la richiesta del client a favore di un piano diverso che comunque rispetti lo scopo delegato..
- **Critical-Sub help**: Il contractor si limita alla sola realizzazione di un sottoscopo modificando comunque l'azione (rispetto a quella delegata) per raggiungere quel sottoscopo.
- **Hyper-critical help**: il contractor raggiunge scopi (del delegante) che però il elegante non ha considerato: facendo ciò, il contractor né implementa il compito dele-

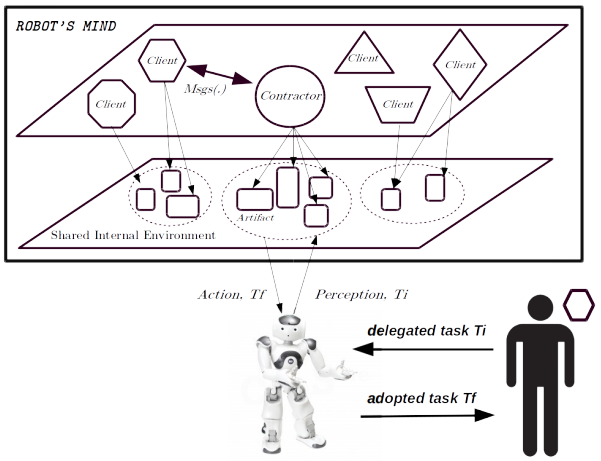

Figura 1: schema del modello computazionale

gato, né raggiunge totalmente gli scopi che gli sono stati delegati.

Il nostro contributo è stato quello di sviluppare un modello computazionale dichiarativo, knowledge-oriented che integri il concetto di cooperazione descritto in [Castelfranchi e Falcone, 1998], come base per un'effettiva interazione tra uomo e robot. Per l'implementazione del modello abbiamo sfruttato le potenzialità del framework *JaCaMo* [Boissier *et al.*, 2013], una piattaforma di multi-agent programming che integra tre diversi livelli di programmazione multi-agent: agent-oriented programming, environment-oriented programming, organization-oriented programming. Ogni livello è associato ad una piattaforma software esistente e ben consolidata, ma utilizzata in modo indipendente.

2 Descrizione del modello cognitivo

La figura 1 fornisce una visione generale del modello. Di fatto, tale modello è un sistema multi-agente che globalmente rappresenta la mente del robot e che è popolato da due categorie di agenti BDI: il *Contractor* e i *Clients*. Nel nostro caso, il contractor è una rappresentazione del robot, mentre i clients sono rappresentazioni di suoi possibili interlocutori. Ogni rappresentazione è descritta da proprie beliefs, goals e piani. Le beliefs sono formule logiche che codificano informazioni percepite dal mondo, attribuite o che provengono dalla comunicazione con altri agenti coinvolti nella interazione. I goals sono stati del mondo che un agente vuole raggiugere, ad esempio un goal delegato dall'esterno. L'agente raggiunge goals eseguendo piani appartenti alla propria libreria, che stabilisce il know-how dell'agente. Ogni agente può operare attraverso un proprio e estendibile ciclo di ragionamento, in base a cui aggiorna le proprie beliefs e raggiunge goals selezionando piani, le cui precondizioni sono compatibili con lo stato corrente della cooperazione. Ogni agente può comunicare con altri agenti trasferendo conoscenza. Infatti, sia il modello del robot che quello dei clients, possono scambiarsi messaggi, percepire e agire in un ambiente condiviso che fornisce strumenti di supporto agli agenti e che, di fatto, incapsula anche le funzionalità del robot. La presenza, nella mente del robot, di un modello di sè stesso e di modelli di altri agenti coinvolti nella interazione, risulta molto potente. Infatti, il robot può adattare il proprio livello di autonomia nella adozione di un compito delegato dall'esterno, scambiandosi messaggi con il modello del client coinvolto nella interazione e decidere sulla base di una conoscenza dello stato corrente dell'ambiente percepito,sulla base della conoscenza del proprio stato interno e sfruttando tecniche di plan recognition e intention recognition per navigare tra i piani e gli scopi attribuiti all'interlocutore. Il modello del robot introduce un ulteriore elemento di descrizione delle capacità del robot reale, dei propri limiti fisici e tecnologici.

3 Esperimenti

Il modello è stato testato in un dominio specifico, sfruttando un robot reale molto diffuso in ambito human-robot interaction: il robot umanoide Nao. Nao ha ricoperto il ruolo di infoPoint assistant della città di Roma. Durante l'interazione con diverse categorie di utenti, il robot è riuscito ad adattare la propria autonomia collaborativa mostrando diversi gradi di adozione del compito delegato dall'utente. Il modello si è dimostrato molto potente nel conferire al robot la capacità di andare oltre alla semplice accettazione della delega risolvendo anche eventuali conflitti di natura collaborativa, dovuti all'iniziativa del robot di andare oltre agli scopi delegati dall'utente.

4 Conclusioni e sviluppi futuri

Questo lavoro conclude la prima fase di un progetto di robotica cognitiva, il cui obiettivo finale sarà quello di sviluppare un modello cognitivo che permetta di rappresentare, nella mente di un robot, lo stato mentale della fiducia [Castelfranchi e Falcone, 2010]. Realizzare meccanismi decisionali che hanno in se il concetto di delega e adozione, secondo la metodologia proposta precedentemente, è il primo passo per implementare sistemi artificiali in grado di instaurare rapporti di fiducia con utenti umani.

Riferimenti bibliografici

[Boissier *et al.*, 2013] Olivier Boissier, Rafael H Bordini, Jomi F Hübner, Alessandro Ricci, e Andrea Santi. Multi-agent oriented programming with jacamo. *Science of Computer Programming*, 78(6):747–761, 2013.

[Castelfranchi e Falcone, 1998] Cristiano Castelfranchi e Rino Falcone. Towards a theory of delegation for agent-based systems. *Robotics and Autonomous Systems*, 24(3-4):141–157, 1998.

[Castelfranchi e Falcone, 2010] Christiano Castelfranchi e Rino Falcone. *Trust theory: A socio-cognitive and computational model*, volume 18. John Wiley & Sons, 2010.

[Kajdocsi e Pozna, 2014] Lószló Kajdocsi e Claudiu Radu Pozna. Review of the most successfully used cognitive architectures in robotics and a proposal for a new model of knowledge acquisition. pages 239–244, 2014.