Un Agente in Grado di Giustificare le Proprie Azioni

Angelo Maria Pio Sabella[1,*], Valeria Seidita[1] and Antonio Chella[1,2]

[1]Universitá degli Studi di Palermo
[2]ICAR-CNR, Palermo, Italy

Abstract
Con questo breve articolo vogliamo presentare il lavoro svolto e le ricerche in atto presso il RoboticsLab dell'Università degli Studi di Palermo. Le tematiche affrontate trattano la realizzazione di un modello che permetta a un agente intelligente di fornire una giustificazione sul proprio operato. Per fare questo sfruttiamo i concetti del paradigma ad agenti BDI e dotiamo l'agente della capacità di illustrare le ragioni del suo agire attraverso quello che viene definito "inner-speech". Dotare un agente della capacità di giustificare le proprie scelte permette di migliorare il livello di fiducia degli esseri umani ed al tempo stesso di fornire gli strumenti per condurre il processo decisionale.

Keywords
Human Agent Interaction, Explainability, Inner Speech, JaCaMo

1. Introduction

L'interesse per i sistemi capaci di adattarsi autonomamente e dotati di autocoscienza è in rapida crescita in questi anni. Dotare robot o agenti di capacità cognitive è certamente la prossima svolta nel campo dell'intelligenza artificiale. Tale obiettivo mira ad una migliore interazione uomo-robot in contesti cooperativi. L'interazione uomo-robot (HRI) è la disciplina che studia come analizzare e sviluppare robot che interagiscono con l'uomo per perseguire un obiettivo comune.

L'interazione è il processo di collaborazione per raggiungere un obiettivo e può essere vista da diversi punti di vista. Essa prevede tra le altre cose: la comunicazione tra i partecipanti, la condivisione e l'aggiornamento delle proprie conoscenze che costituiscono sfide significative dal punto di vista della ricerca. Inoltre, in quanto esseri umani, siamo portati a lavorare meglio in gruppo con persone di cui ci fidiamo: questo pone un ulteriore livello di difficoltà. Oltre a questo, la fiducia è anche un parametro che influenza le decisioni relative alle attività da compiere e/o delegare. L'obiettivo finale delle ricerche in corso è quello di implementare le interazioni in team composti sia da umani che da robot in modo che la collaborazione risulti il più efficiente e affidabile possibile.

Obiettivo principale del nostro lavoro è un modello cognitivo, e la sua relativa implementazione, per dotare un agente intelligente di un modulo che consenta di giustificare le scelte relative alle azioni compiute per il raggiungimento del compito designato (o per gli step intermedi). Questo ci permette di ottenere un agente in grado di esternare in qualche modo il proprio processo di ragionamento e di fornire una spiegazione all'essere umano con cui interagisce.

Ottenere una spiegazione dell'operato da parte dell'agente ha un impatto positivo durante lo svolgimento delle operazioni e in termini di fiducia, in quanto l'umano potrà comprendere i comportamenti adottati. La capacità degli agenti di poter giustificare le loro azioni mira a creare degli agenti sempre più affidabili e credibili.

2. Modello proposto

L'idea è basata sulla creazione di un modello cognitivo e della corrispondente architettura per un agente intelligente, i cui moduli consentano di strutturare il processo decisionale in modo da tener conto anche degli stati interni. A livello implementativo, sono state esaminate le prestazioni del paradigma ad agenti BDI (Belief-Desire-Intention) [1, 2] tramite l'utilizzo del framework JaCaMo [3, 4, 5]. JaCaMo permette di integrate le tre dimensioni che compongono un sistema multi-agente (MAS): agenti, ambiente e organizzazione . Jason [6] fornisce un linguaggio basato sul modello BDI, nello specifico AgentSpeak(L) [7], per la definizione degli agenti, CArtAgO [8, 9] permette di programmare l'ambiente in Java, e Moise infine è utilizzato per definire il livello organizzativo.

JaCaMo fornisce l'astrazione necessaria per poter definire tutti i componenti di un sistema multi-agente tramite moduli separati, ma integrati fra loro.

Partendo dal modello ottenuto nei lavori precedenti [10, 11, 12, 5], il passaggio successivo prevede l'implementazione di un modulo che fornisca all'agente la capacità di ragionare ad alta voce, cioè di esternare il proprio ragionamento rendendo comprensibili le motivazioni alla base delle proprie azioni agli esseri umani. A tal scopo, è stata

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
angelomariapio.sabella@unipa.it (A. M. P. Sabella); valeria.seidita@unipa.it (V. Seidita); antonio.chella@unipa.it (A. Chella)
0000-0002-0601-6914 (V. Seidita); 0000-0002-8625-708X (A. Chella)

proposta una possibile estensione del ciclo di ragionamento dell'interprete Jason, basato sul pratical reasoning, per adattarlo ai nostri scopi [5].

L'idea è stata integrare all'interno del ragionamento una serie di funzioni che permettono all'agente di poter selezionare l'azione in base alle proprie credenze e alle proprie capacità, valutare quest'ultima e fornire una giustificazione delle propria scelta. Un primo modello semplificato tiene conto degli elementi chiave necessari a dotare l'agente delle funzionalità richieste: la fase di percezione del mondo esterno e di quello interno, il processo di deliberazione e la fase di azione. In particolare, l'azione, opportunamente selezionata come spiegato in precedenza, necessita di essere valutata prima di essere eseguita. È in questa fase che l'agente è in grado di fornire una giustificazione prima di procedere con l'esecuzione. Grazie alla presenza del modulo di inner speech, l'agente è in grado di esternare il proprio discorso interiore relativo al processo che porta a tale selezione, alla valutazione e alla successiva esecuzione. Dotare l'agente di tale capacità significa rendere il comportamento di quest'ultimo più spiegabile, comprendendone le motivazioni di fondo.

3. La Giustificazione tramite l'Inner Speech e gli Speech Act

Per introdurre tali funzionalità si è pensato di colmare il divario tra il modello proposto e la sua implementazione legando il concetto di inner-speech [13, 14, 15, 16] con quello degli speech act [17].

Una definizione esatta di inner speech è difficile da elaborare. Alcuni autori definiscono il discorso interiore come l'esperienza personale del linguaggio in relazione alle proprie azioni, ai propri stati mentali e alle esperienze. Inoltre, è spesso associato a concetti quali l'autoconsapevolezza e l'autocoscienza [18, 19]. La prima è "la capacità di diventare l'oggetto della propria attenzione" [20] prendendo in esame tre elementi: l'ambiente sociale, il mondo fisico e sé stessi. La seconda comprende la capacità di focalizzare l'attenzione su un preciso stato mentale.

Nei sistemi multi-agente la comunicazione si basa sulla teoria degli speech act. In particolare, in Jason questa è supportata dal linguaggio di comunicazione tra agenti noto come Knowledge Query and Manipulation Language (KQML) [21, 22, 23].

Basandosi su tale meccanismo di comunicazione e sul fatto che l'inner speech è una forma di dialogo con sé stessi, abbiamo pensato di utilizzare gli speech act come strumento per rendere l'agente in grado di inviare dei messaggi a sé stesso, realizzando così, per ora in modo semplice, il discorso interiore. Questo fornisce un mezzo per esternare il ragionamento dell'agente in merito agli eventi che percepisce, aggiornare le proprie credenze e dare una spiegazione di ciò che sta elaborando.

4. Prima implementazione e obiettivi futuri

Una prima implementazione del modello proposto è stata realizzata sulla base di uno scenario collaborativo già presente nei lavori passati: l'apparecchiamento della tavola. Il motivo di tale scelta è molto semplice: si è pensato di riutilizzare tale scenario così da rendere eventualmente possibile, in futuro, il confronto tra il modello che utilizza il paradigma ad agenti BDI con i risultati degli esperimenti già svolti in precendenza e che prevedono l'uso di ACT-R [24] come tecnologia.

Ulteriore passo in avanti che stiamo portando avanti sfrutta un'altra particolare componente di JaCaMo. Grazie a questa possiamo introdurre un quarto elemento fondamentale nei sistemi multi-agente rendendolo programmabile in maniera indipendente dai tre domini già citati, ma allo stesso tempo legandolo ad essi tramite opportuni collegamenti. Possiamo definire un protocollo di interazione [25] tra i partecipanti coinvolti. Quindi la possibilità di esplicitare come svolgere un certo compito per raggiungere l'obiettivo finale, chi e cosa sono necessari e le diverse transizioni di stato durante il processo.

Inoltre, pensiamo di semplificare il ragionamento dell'agente per quanto riguarda l'inner speech. Poiché stiamo trattando una forma di discorso interiore, quindi con sé stessi, si può pensare di definire esattamente quali sono gli step che compongono questa interazione, fatta di un singolo partecipante: l'agente stesso. Abbiamo quindi la possibilità di formalizzare il concetto di inner speech e di giustificazione per un agente, senza dover complicare ulteriormente l'interprete Jason.

5. Conclusioni

L'idea alla base delle ricerche condotte è quella di riuscire a dotare un agente di capacità linguistiche interne al fine di poter esternare il proprio discorso interiore e rendere l'interazione con l'essere umano più affidabile e comprensibile. Le scelte, dal punto di vista tecnologico, sono ricadute sul paradigma ad agenti BDI e sull'utilizzo di JaCaMo, grazie al pieno supporto di tale paradigma. Inoltre, risulta semplice individuare un parallelismo tra il modello proposto e il ciclo di ragionamento implementato dall'interprete Jason. Per poter rappresentare uno scenario completo e complesso, però, è necessario progettare con cura il sistema, in quanto bisogna individuare con attenzione tutti gli elementi che andranno a comporre il sistema e soprattutto la conoscenza necesaria all'agente per operare correttamente. Questi problemi saranno parte dei nostri lavori futuri.

6. Ringraziamenti

La ricerca descritta nel presente articolo è parzialmente supportata dal progetto AFOSR RESPECT FA9550-19-1-7025.

References

[1] M. Georgeff, A. Rao, Rational software agents: from theory to practice, in: Agent technology, Springer, 1998, pp. 139–160.

[2] A. S. Rao, M. P. Georgeff, et al., BDI agents: from theory to practice., in: ICMAS, volume 95, 1995, pp. 312–319.

[3] O. Boissier, R. Bordini, J. Hübner, A. Ricci, A. Santi, Multi-agent oriented programming with jacamo, Science of Computer Programming 78 (2013) 747–761.

[4] O. Boissier, R. Bordini, J. Hubner, A. Ricci, Multi-agent oriented programming: programming multi-agent systems using JaCaMo, MIT Press, 2020.

[5] V. Seidita, F. Lanza, A. Sabella, A. Chella, Can agents talk about what they are doing? a proposal with jason and speech acts, volume 3261, 2022, pp. 17–29. Cited by: 0.

[6] R. Bordini, J. Hübner, BDI agent programming in AgentSpeak using jason, in: International Workshop on Computational Logic in Multi-Agent Systems, Springer, 2005, pp. 143–164.

[7] A. S. Rao, Agentspeak (l): BDI agents speak out in a logical computable language, in: European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Springer, 1996, pp. 42–55.

[8] A. Omicini, A. Ricci, M. Viroli, Artifacts in the a&a meta-model for multi-agent systems, Autonomous agents and multi-agent systems 17 (2008) 432–456.

[9] A. Ricci, M. Viroli, A. Omicini, Cartago: A framework for prototyping artifact-based environments in mas, in: International Workshop on Environments for Multi-Agent Systems, Springer, 2006, pp. 67–86.

[10] V. Seidita, C. Diliberto, P. Zanardi, A. Chella, F. Lanza, Inside the robot's mind during human-robot interaction, in: 7th International Workshop on Artificial Intelligence and Cognition, AIC 2019, volume 2483, CEUR-WS, 2019, pp. 54–67.

[11] A. Chella, F. Lanza, V. Seidita, A cognitive architecture for human-robot teaming interaction, in: Proceedings of the 6th International Workshop on Artificial Intelligence and Cognition, Palermo, 2018.

[12] C. Castelfranchi, A. Chella, R. Falcone, F. Lanza, V. Seidita, Endowing robots with self-modeling abilities for trustful human-robot interactions, in: 20th Workshop" From Objects to Agents", WOA 2019, volume 2404, CEUR-WS, 2019, pp. 22–28.

[13] L. E. Berk, R. Garvin, Development of private speech among low-income appalachian children., Developmental psychology 20 (1984) 271.

[14] A. Morin, Possible links between self-awareness and inner speech theoretical background, underlying mechanisms, and empirical evidence, Journal of Consciousness Studies 12 (2005) 115–134.

[15] L. Vygotsky, M. Cole, Mind in society: Development of higher psychological processes, Harvard university press, 1978.

[16] A. Winsler, C. Fernyhough, I. Montero, Private speech, executive functioning, and the development of verbal self-regulation., Cambridge University Press, 2009.

[17] J. Searle, J. R. Searle, Speech acts: An essay in the philosophy of language, volume 626, Cambridge university press, 1969.

[18] P. Silvia, T. Duval, Objective self-awareness theory: Recent progress and enduring problems, Personality and social psychology review 5 (2001) 230–241.

[19] A. Morin, A neurocognitive and socioecological model of self-awareness, Genetic, social, and general psychology monographs 130 (2004) 197–224.

[20] S. Duval, R. Wicklund, A theory of objective self awareness. (1972).

[21] T. Finin, R. Fritzson, D. McKay, R. McEntire, Kqml as an agent communication language, in: Proceedings of the third international conference on Information and knowledge management, 1994, pp. 456–463.

[22] T. Finin, R. Fritzson, D. P. McKay, R. McEntire, et al., Kqml-a language and protocol for knowledge and information exchange, in: 13th Int. Distributed Artificial Intelligence Workshop, 1994, pp. 93–103.

[23] Y. Labrou, T. Finin, A semantics approach for kqml—a general purpose communication language for software agents, in: Proceedings of the third international conference on Information and knowledge management, 1994, pp. 447–455.

[24] J. Anderson, M. Matessa, C. Lebiere, Act-r: A theory of higher level cognition and its relation to visual attention, Human–Computer Interaction 12 (1997) 439–462.

[25] M. R. Zatelli, J. F. Hübner, The interaction as an integration component for the jacamo platform, in: Engineering Multi-Agent Systems: Second International Workshop, EMAS 2014, Paris, France, May 5-6, 2014, Revised Selected Papers 2, Springer, 2014, pp. 431–450.

A. Risorse Online

È disponibile online una repository GitHub contenente un primo tentativo di implementazione di una parte del modello descritto nell'articolo presentata al 23rd Workshop "From Objects to Agents" (WOA22).

- GitHub