Il Ruolo delle Scienze Cognitive nell'Intelligenza Artificiale del Futuro

Antonio Lieto*, Fabio Paglieri **

* Università di Torino, Dipartimento di Informatica - ICAR-CNR, Palermo
** Istituto di Scienze e Tecnologie della Cognizione - CNR, Roma
antonio.lieto@unito.it, fabio.paglieri@istc.cnr.it

Abstract

Questo contributo si propone di fornire uno spunto di riflessione, e una breve panoramica storica, sul ruolo che le scienze cognitive hanno giocato, e possono ancora giocare, nello sviluppo dei sistemi intelligenti di nuova generazione. Illustra, inoltre, le attività recenti che l'AISC (Associazione Italiana di Scienze Cognitive, di cui gli autori sono attualmente Vice-Presidente e Presidente) sta portando avanti per lo sviluppo di linee di ricerca nell'ambito dei sistemi artificiali di inspirazione cognitiva.

1 Introduzione

La ricerca nell'ambito dell'Intelligenza Artificiale (IA) si è storicamente basata, soprattutto agli albori della disciplina, su una forte collaborazione tra informatici, ingegneri, biologi, filosofi e psicologi impegnati a lavorare in quel campo di ricerca interdisciplinare che oggi prende il nome di "Scienze Cognitive".

Tale collaborazione, favorita soprattutto dalla forte influenza esercitata dell'approccio cibernetico allo studio di sistemi naturali e artificiali [Cordeschi, 2002], ha prodotto, nel corso degli anni, lo sviluppo di fruttuose linee di ricerca in bionica, robotica, sistemi biologicamente inspirati e, più in generale, nell'ambito dei sistemi artificiali di inspirazione cognitiva e della scienza dei sistemi (o teoria dei sistemi).

Dopo decenni di mutua e pioneristica collaborazione, tuttavia, a partire dalla metà degli anni '80 del secolo scorso, l'Intelligenza Artificiale e le Scienze Cognitive hanno iniziato a produrre diverse sottodiscipline, ciascuna caratterizzata da obiettivi propri e da propri metodologie di ricerca e di valuazione [Langley, 2012].

Questa frammentazione, se da un lato ha facilitato lo sviluppo di sistemi di IA in grado di riprodurre (e in alcuni casi di superare) le competenze umane in ambiti ristretti (ad es. negli scacchi, nella dimostrazione di teoremi, in giochi come Jeopardy [Ferrucci, 2012] o Go [Silver *et al.*, 2016]), d'altra parte si è focalizzata su un approccio *divide et impera* di tipo verticale che ha inibito in modo significativo la collaborazione orizzontale tra diverse discipline (*cross-field collaboration*) e lo sforzo scientifico finalizzato ad investigare obiettivi più generali quali, ad esempio: i) cosa si intende per comportamento intelligente (in sistemi autonomi naturali e artificiali) e ii) come, e fino a che punto, è possibile progettare e costruire artefatti intelligenti prendendo spunto dalle euristiche esibite da umani - o da altri animali - nel mondo "naturale".

In anni recenti, tuttavia, l'area dei sistemi artificiali di ispirazione cognitiva ha attratto un rinnovato interesse sia in ambito accademico che industriale [Lieto e Radicioni, 2016]. La rinnovata presa di coscienza della necessità di realizzare sistemi in grado di operare, con le stesse "primitive", in contesti diversi è infatti alla base della constatazione che il gap tra sistemi naturali e artificiali è, ancora oggi, enorme (specialmente se si considerano domini realistici).

2 Il ruolo dell'approccio cognitivo in IA

Il successo recente dei sistemi di IA si deve principalmente al paradigma dell'apprendimento automatico basato su reti neurali profonde (Deep Learning) e all'utilizzo integrato di diversi modelli probalistici (si pensi, ad esempio, al sistema Watson di IBM). Le perfomance super-umane, e *domanin-specific*, di alcuni sistemi basati su tali approcci, tuttavia, sono principalmente dovute alla aumentata potenza di calcolo dei computer odierni e all'enorme disponibilità attuale di dati (i cosiddetti *big data*) piuttosto che all'invenzione di modelli di calcolo nuovi o innovativi rispetto a quelli già noti e sviluppati nell'ultimo trentennio.

Allo stesso tempo, i sistemi basati sugli approcci di IA attualmente più in voga, si sono resi protagonisti - oltre che dei già menzionati successi - di altrettanto eclatanti errori che ne hanno messo in luce alcuni punti deboli. Ad esempio, il sistema Watson ha mostrato notevoli limiti nel rispondere a domande apparentemente semplici (almeno per gli esseri umani) ma che implicano la capacità di utilizzare forme di ragionamento di senso comune [Davis e Marcus, 2015]. Anche il sistema di deep learning alla base di Alpha Go (inizialmente utilizzato per compiti di categorizzazione automatica di immagini) è salito alla ribalta della cronaca per aver classificato una coppia di afro-americani come dei gorilla (provocando, come si può immaginare, un dibattito molto acceso circa i problemi etici dell'IA scaturiti dall'utilizzo di dataset "con bias")[1]. Ancora, i sistemi attuali di deep learning si sono mostrati poco robusti ad attacchi generati dalle cosiddette Ad-

[1] https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/

versarial Networks (reti neurali addestrate per imparare quali elementi dell'input modificare per determinare output diversi). In alcuni casi, infatti, i sistemi *deep* sono stati "ingannati" da manipolazioni riguardanti un pugno di pixel (da 1 a 5) all'interno di una immagine [Su *et al.*, 2019].

Tutti questi errori (che ovviamente non hanno alcuna pretesa di rappresentatività o esaustività) hanno un elemento in comune: non sarebbero mai stati commessi da alcun essere umano. Nello specifico: tutti i compiti che sembrano particolarmente "semplici" da eseguire per gli esseri umani (e, in alcuni casi, per altre specie animali) rappresentano degli ostacoli molto difficili da superare per gli attuali approcci di IA. A nostro avviso sono proprio questi i tipi di problemi in cui si evidenzia il rinnovato bisogno di una integrazione dell' approccio euristico di ispirazione cognitiva nell'ambito delle fasi di progettazione e implementazione di sistemi intelligenti.

Tali problemi riguardano, a nostro avviso, almeno le seguenti aree (la lista non è esaustiva):

- Ragionamento di senso comune in compiti spaziali, *action-oriented*, in ambienti dinamici [Chella *et al.*, 2000] e in task di categorizzazione [Bara *et al.*, 2001], [Lieto *et al.*, 2017]
- Ragionamento analogico [Bianchini, 2016] e metaforico [Lieto e Pozzato, 2019]
- Apprendimento da pochi esempi o da singoli esempi (few shot learning e one shot learning) [Gagliardi, 2008]
- Trasferimento dell'apprendimento in contesti multi-dominio [Mantovani e Castelnuovo, 2003]
- Creatività computazionale [Augello *et al.*, 2016], [Lieto e Pozzato, 2019]
- Elaborazione del linguaggio naturale e comprensione narrativa [Airenti *et al.*, 1993; Lenci, 2008; Cangelosi e Parisi, 2012; Lieto *et al.*, 2015]
- Integrazione euristica di percezioni multimodali [Pezzulo *et al.*, 2013]
- Integrazione robusta di meccanismi che riguardano le attività di pianificazione, monitoraggio, azione e ragionamento guidato da obiettivi (*goal-reasoning*) [Castelfranchi, 1998], [Falcone e Castelfranchi, 2001], [Conte *et al.*, 2016] [Lieto *et al.*, 2018b]
- Modellazione di emozioni [Plebe, 2016] e di meccanismi di cognizione sociale e multi-agente [Castelfranchi e Falcone, 1998], [Paglieri *et al.*, 2014] [Lieto *et al.*, 2018a]
- Robotica cognitiva e sociale, [Miglino *et al.*, 1995; **?**],
- Generazione di spiegazioni comprensibili ad utenti umani di decisioni algoritmiche (Explainable AI) [Colla *et al.*, 2018]

Come evidenzia questa lista, i problemi menzionati in ciascuna area coinvolgono sia capacità eminentemente percettive che capacità cognitive di alto livello (modellate, tipicamente, con approcci connessionisti, nel primo caso, e simbolici o ibridi nel secondo). Come ha dimostrato la storia delle Scienze Cognitive negli ultimi decenni, dunque, l'agenda cognitiva è compatibile con attività di modellazione computazionale sia di tipo neurale che di tipo simbolico.

3 L'Associazione Italiana di Scienze Cognitive

L'AISC (Associazione Italiana di Scienze Cognitive) è da sempre impegnata in attività di ricerca nell'ambito dell'IA di inspirazione cognitiva. In tutti gli ambiti precedentemente indicati, e in molti altri, la comunità nazionale di scienziati cognitivi computazionali è, infatti, fortemente impegnata e inserita nei principali contesti internazionali. Per favorire ulteriormente l'integrazione e il dialogo tra IA e Scienze Cognitive, l'associazione ha di recente avviato una attività di collaborazione e dialogo con l'AI*IA (Associazione Italiana di Intelligenza Artificiale) tramite l'organizzazione del primo panel congiunto tra le due associazioni svoltosi presso la conferenza internazionale AI*IA 2017 a Bari. Ha, inoltre, sponsorizzato iniziative come la Advanced School in AI (https://as-ai.org/) e la serie di workshop internazionali AIC (su "Artificial Intelligence and Cognition", https://dblp.org/db/conf/aic/index). L'ultimo convegno dell'associazione (AISC 2018), infine, è stato focalizzato sul tema "The new era of Artificial Intelligence: a cognitive perspective". Ciò, a testimonianza del fatto che il forte interesse dell'AISC per il tema dell'IA non riguarda solo il passato ma anche gli sviluppi futuri del settore. Tale base, scientifica e culturale, può rappresentare, a nostro avviso, il punto di partenza per lo sviluppo di future collaborazioni con il CINI AIIS Lab.

Ringraziamenti

Il contenuto di questo intervento ha beneficiato delle riflessioni emerse nell'ambito del panel "Can AI and Cognitive Science still live together happily ever after?" (link: http://aiia2017.di.uniba.it/index.php/joint-panel-aiia-and-aisc/) organizzato presso la XV conferenza internazionale AI*IA 2017, Bari. Ringraziamo in particolare: Amedeo Cesta, Antonio Chella, Oliviero Stock e Giuseppe Trautteur per i loro interventi.

Riferimenti bibliografici

[Airenti *et al.*, 1993] Gabriella Airenti, Bruno G Bara, e Marco Colombetti. Conversation and behavior games in the pragmatics of dialogue. *Cognitive science*, 17(2):197–256, 1993.

[Augello *et al.*, 2016] Agnese Augello, Ignazio Infantino, Antonio Lieto, Giovanni Pilato, Riccardo Rizzo, e Filippo Vella. Artwork creation by a cognitive architecture integrating computational creativity and dual process approaches. *Biologically inspired cognitive architectures*, 15:74–86, 2016.

[Bara *et al.*, 2001] Bruno G Bara, Monica Bucciarelli, e Vincenzo Lombardo. Model theory of deduction: A unified computational approach. *Cognitive Science*, 25(6):839–901, 2001.

[Bianchini, 2016] Francesco Bianchini. Analogy as categorization: A support for model-based reasoning. In

Model-Based Reasoning in Science and Technology, pages 239–256. Springer, 2016.

[Cangelosi e Parisi, 2012] Angelo Cangelosi e Domenico Parisi. *Simulating the evolution of language*. Springer Science & Business Media, 2012.

[Castelfranchi e Falcone, 1998] Cristiano Castelfranchi e Rino Falcone. Principles of trust for mas: Cognitive anatomy, social importance, and quantification. In *Proceedings International Conference on Multi Agent Systems (Cat. No. 98EX160)*, pages 72–79. IEEE, 1998.

[Castelfranchi, 1998] Cristiano Castelfranchi. Modelling social action for ai agents. *Artificial intelligence*, 103(1-2):157–182, 1998.

[Chella *et al.*, 2000] Antonio Chella, Marcello Frixione, e Salvatore Gaglio. Understanding dynamic scenes. *Artificial intelligence*, 123(1-2):89–132, 2000.

[Colla *et al.*, 2018] Davide Colla, Enrico Mensa, Daniele P Radicioni, e Antonio Lieto. Tell me why: Computational explanation of conceptual similarity judgments. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 74–85. Springer, 2018.

[Conte *et al.*, 2016] Rosaria Conte, Cristiano Castelfranchi, et al. *Cognitive and social action*. Garland Science, 2016.

[Cordeschi, 2002] Roberto Cordeschi. *The discovery of the artificial: Behavior, mind and machines before and beyond cybernetics*, volume 28. Springer Science & Business Media, 2002.

[Davis e Marcus, 2015] Ernest Davis e Gary Marcus. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9):92–103, 2015.

[Falcone e Castelfranchi, 2001] Rino Falcone e Cristiano Castelfranchi. Social trust: A cognitive approach. In *Trust and deception in virtual societies*, pages 55–90. Springer, 2001.

[Ferrucci, 2012] David A Ferrucci. Introduction to "this is watson". *IBM Journal of Research and Development*, 56(3.4):1–1, 2012.

[Gagliardi, 2008] Francesco Gagliardi. A prototype-exemplars hybrid cognitive model of "phenomenon of typicality" in categorization: A case study in biological classification. In *Proc. 30th Annual Conf. of the Cognitive Science Society, Austin, TX*, pages 1176–1181, 2008.

[Langley, 2012] Pat Langley. The cognitive systems paradigm. *Adv. in Cognitive Systems*, 1:3–13, 2012.

[Lenci, 2008] Alessandro Lenci. Distributional semantics in linguistic and cognitive research. *Italian journal of linguistics*, 20(1):1–31, 2008.

[Lieto *et al.*, 2015] Antonio Lieto, Daniele P Radicioni, e Valentina Rho. A common-sense conceptual categorization system integrating heterogeneous proxytypes and the dual process of reasoning. In *Proceedings of IJCAI 2015*, pages 875–881, 2015.

[Lieto *et al.*, 2017] Antonio Lieto, Daniele P Radicioni, e Valentina Rho. Dual peccs: a cognitive system for conceptual representation and categorization. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(2):433–452, 2017.

[Lieto *et al.*, 2018a] Antonio Lieto, William G Kennedy, Christian Lebiere, Oscar Romero, Niels Taatgen, e Robert West. Higher-level knowledge, rational and social levels constraints of the common model of the mind. *Procedia Computer Science*, 2018.

[Lieto *et al.*, 2018b] Antonio Lieto, Christian Lebiere, e Alessandro Oltramari. The knowledge level in cognitive architectures: Current limitations and possible developments. *Cognitive Systems Research*, 48:39–55, 2018.

[Lieto e Pozzato, 2019] Antonio Lieto e Gian Luca Pozzato. A description logic framework for commonsense conceptual combination integrating typicality, probabilities and cognitive heuristics. *arXiv preprint arXiv:1811.02366*, 2019.

[Lieto e Radicioni, 2016] Antonio Lieto e Daniele P Radicioni. From human to artificial cognition and back: New perspectives on cognitively inspired ai systems. *Cognitive Systems Research*, 2016.

[Mantovani e Castelnuovo, 2003] Fabrizia Mantovani e Gianluca Castelnuovo. The sense of presence in virtual training: enhancing skills acquisition and transfer of knowledge through learning experience in virtual environments. 2003.

[Miglino *et al.*, 1995] Orazio Miglino, Henrik Hautop Lund, e Stefano Nolfi. Evolving mobile robots in simulated and real environments. *Artificial life*, 2(4):417–434, 1995.

[Paglieri *et al.*, 2014] Fabio Paglieri, Cristiano Castelfranchi, Célia da Costa Pereira, Rino Falcone, Andrea Tettamanzi, e Serena Villata. Trusting the messenger because of the message: feedback dynamics from information quality to source evaluation. *Computational and Mathematical Organization Theory*, 20(2):176–194, 2014.

[Pezzulo *et al.*, 2013] Giovanni Pezzulo, Lawrence W Barsalou, Angelo Cangelosi, Martin H Fischer, Ken McRae, e Michael Spivey. Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Frontiers in psychology*, 3:612, 2013.

[Plebe, 2016] Alessio Plebe. What is 'wrong'in a neural model. *Cognitive Systems Research*, 39:4–14, 2016.

[Silver *et al.*, 2016] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484, 2016.

[Su *et al.*, 2019] Jiawei Su, Danilo Vasconcellos Vargas, e Kouichi Sakurai. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 2019.