Strumenti di logica computazionale per estrazione e iniezione di conoscenza simbolica da e verso predittori subsimbolici

1 Progetto di ricerca

La presente attività di ricerca si colloca nell'ambito del più ampio progetto CHIST-ERA "EXPECTATION" (G.A. CHIST-ERA-19-XAI-005) [2], il quale ha come obiettivo la progettazione e prototipazione di un sistema di raccomandazioni che integri AI simbolica e sub-simbolica mediante un approccio multi-agente, al fine di fornire raccomandazioni efficaci e spiegazioni personalizzate. In particolare, EXPECTATION prevede che il normale approccio data-driven – in cui le raccomandazioni sono costruite per mezzo di dati numerici elaborati tramite apprendimento automatico – sia complementato dall'impiego tecniche di AI simbolica al fine di migliorare la comprensibilità e malleabilità delle spiegazioni. Più in dettaglio, nel design di base di EXPECTATION, le fasi di approvvigionamento ed analisi dei dati sono incapsulate all'interno di agenti intelligenti, capaci anche e soprattutto di *estrarre* conoscenza simbolica – e in particolare regole logiche – dai predittori subsimbolici usati in fase di apprendimento automatico, mediante appositi algoritmi (p.e. [3]). Della presenza di molteplici agenti siffatti, i sistemi di raccomandazioni basati sull'approccio EXPECTATION fanno il loro maggiore punto di forza. I vari agenti possono avere accesso a dati diversi e impiegare differenti risorse computazionali per la loro elaborazione, dando quidi spazio ad una moltitudine di possibili raccomandazioni e spiegazioni che vanno quindi integrate e opportunamente presentate agli utenti, mediante elaborazioni simboliche. Le conoscenza simbolica estratta da dati e predittori gioca pertanto un ruolo fondamentale sul fronte dell'integrazione. Essa è usata per costruire raccomandazioni e spiegazioni personalizzate mediante l'impiego di: (i) programmazione logica classica e le sue varianti (in particolare probabilistica, induttiva, od abduttiva), per la costruzione di raccomandazioni efficaci, (ii) tecniche di argomentazione e negoziazione per la costruzione di spiegazioni interattive.

Infine, punto cardine di EXPECTATION è la consapevolezza che la spiegazione automatica ed interattiva supporta non soltanto l'utente umano ma anche l'agente software con limitate capacità di accesso ai dati od alle risorse di elaborazione. Quest'ultimo potrebbe complementare le proprie lacune mediante l'interazione con altri agenti e lo scambio di informazione simbolica aggregata. Gioca quindi un ruolo fondamentale la possibilità

di *iniettare* conoscenza simbolica in sistemi sub-simbolici, permettendo ad ogni agente di aggiornare i propri predittori alla luce di nuova e migliore informazione proveniente dall'esterno.

Ricapitolando, EXPECTATION mira a combinare una serie di contributi a cavallo tra l'AI simbolica e sub-simbolica, tutti fondamentalmente intersecati con la logica computazionale:

1. tecniche di estrazione di conoscenza simbolica (regole) da predittori sub-simbolici,
2. tecniche di iniezione di conoscenza simbolica nei suddetti predittori,
3. tecniche di ragionamento automatico deduttivo, probabilistico, od abudittivo,
4. tecniche di argomentazione, ragionamento ritrattabile (*defeasible*), e negoziazione automatica.

Affinché tali tecniche siano agevolmente trattabili ed armonizzabili è necessario tuttavia un supporto tecnologico adeguato. Mentre per l'AI sub-simbolica e per l'apprendimento automatico sono disponibili oggi una varietà di librerie e strumenti *open-source* molto maturi, sviluppati da gruppi di ricerca sparsi per il globo – spesso con il supporto di grossi attori industriali –, la stessa cosa non si può dire dell'AI simbolica, che consta ad oggi poche soluzioni monolitiche e sufficientemente mature per la programmazione logica classica, e molte soluzioni prototipali che intersecano i più disparati meandri della frontiera della ricerca. Se si considera l'insieme di tecnologie nel suo complesso, ci si scontra poi con un altro problema, ossia la grande eterogeneità delle piattaforme impiegate e la conseguente barriera all'interoperabilità. In particolare si nota come la maggior parte delle tecnologie a supporto dell'AI sub-simbolica siano oggi appannaggio della piattaforma Python, mentre la maggior parte delle soluzioni simboliche siano basate su codice nativo (C, C++). È presente poi una larga varietà di soluzioni software per la ricerca nell'ambito dei sistemi multi-agente che sono invece vincolati alla piattaforma JVM [1].

Per ovviare alle suddette problematiche relative all'interoperabilità e alla scarsa ri-usabilità delle soluzioni software per l'AI simbolica, recentemente, è stato avviato lo sviluppo di 2P-KT¹, una libreria software che mira a divenire un ecosistema aperto, multi-piattaforma e multi-paradigma per l'AI simbolica. La presente attività di ricerca si prefigge pertanto l'obiettivo di estendere il design e lo sviluppo di 2P-KT nelle direzioni utili all'articolazione del progetto EXPECTATION.

2 Piano d'attività

L'attività di ricerca prevede lo sviluppo, la documentazione, e la pubblicazione di software *general-purpose* ed *open-source* a supporto di 4 principali aree tematiche: (i) l'estrazione e (ii) l'iniezione di conoscenza simbolica da/verso predittori subsimbolici, (iii) il

¹<https://github.com/tuProlog/2p-kt>

ragionamento automatico deduttivo, probabilistico, induttivo, od abudittivo, e *(iv)* il ragionamento ritrattabile, con particolare attenzione all'arogmentazione. I prodotti dell'attività comprenderanno pertanto artefatti software, ed i rispettivi artefatti documentali e pubblicazioni scientifiche.

Caratteristiche essenziali delle soluzioni prodotte saranno l'aderenza allo stato stato dell'arte dei rispettivi ambiti e l'interoperabilità interna ed esterna. Per interoperabilità interna si intende la possibilità di interconnettere e combinare le soluzioni prodotte nell'arco dell'attività stessa—ad esempio permettendo il ragionamento automatico a partire da basi di conoscenza estratte da predittori sub-symbolici, o l'iniezione di informazioni simboliche eventualmente indotte o abdotte. Per interoperabilità esterna si intende invece la possibilità per le soluzioni prodotte di interfacciarsi agilmente con i principali *framework* per l'AI sub-simbolica oggi disponibili—da identificarsi nelle fasi preliminari dell'attività di ricerca.

Per ognuna delle aree tematiche suddette, il piano di lavoro è poi sostanzialmente analogo:

1. recognizione dello stato dell'arte ed identificazione delle tecniche più promettenti,
2. progettazione di un'estensione apposita per l'ecosistema 2P-KT riguardante le tecniche individuate,
3. stesura di test automatici, unitari e di sistema, possibilmente multi-piattaforma, che misurino il corretto funzionamento della soluzione da svilupparsi,
4. sviluppo effettivo della soluzione, coerentemente con la letteratura scientifica e le soluzioni già presenti in 2P-KT,
5. produzione dei relativi artefatti documentali,
6. pubblicazione degli artefatti software sugli appositi repository di software open source (Maven Central, PyPi, NPM, etc.),
7. stesura dei relativi articoli scientifici da sottoporre presso le più prestigiose riviste del settore.

Le attività relative alle aree tematiche possono procedere in serie o in parallelo, coerentemente alle necessità e priorità dettate dal progetto EXPECTATION.

Riferimenti bibliografici

- [1] Roberta Calegari, Giovanni Ciatto, Viviana Mascardi, and Andrea Omicini. Logic-based technologies for multi-agent systems: A systematic literature review. *Autonomous Agents and Multi-Agent Systems*, 35(1):1:1–1:67, 2021. Collection “Current Trends in Research on Software Agents and Agent-Based Software Development”.

- [2] Davide Calvaresi, Giovanni Ciatto, Amro Najjar, Reyhan Aydoğan, Leon Van der Torre, Andrea Omicini, and Michael I. Schumacher. EXPECTATION: Personalized explainable artificial intelligence for decentralized agents with heterogeneous knowledge. In Davide Calvaresi, Amro Najjar, Michael Winikoff, and Kary Främling, editors, *Explainable and Transparent AI and Multi-Agent Systems. Third International Workshop, EXTRAAMAS 2021, Virtual Event, May 3–7, 2021, Revised Selected Papers*, volume 12688 of *Lecture Notes in Computer Science*, pages 331–343. Springer Nature, Basel, Switzerland, 2021.
- [3] Federico Sabbatini, Giovanni Ciatto, and Andrea Omicini. GridEx: An algorithm for knowledge extraction from black-box regressors. In Davide Calvaresi, Amro Najjar, Michael Winikoff, and Kary Främling, editors, *Explainable and Transparent AI and Multi-Agent Systems. Third International Workshop, EXTRAAMAS 2021, Virtual Event, May 3–7, 2021, Revised Selected Papers*, volume 12688 of *Lecture Notes in Computer Science*, pages 18–38. Springer Nature, Basel, Switzerland, 2021.