Toulouse, le 30 janvier 2020

PhD thesis proposal

Planning and explanation for artificial intelligent agents

Position at: Institut de Recherche en Informatique de Toulouse (IRIT)
Université de Toulouse, CNRS, INP, UT3
118, route de Narbonne, 31062 Toulouse cedex, France
Department: Artificial Intelligence
Supervisors : Frédéric Maris and Dominique Longin

1 Context and scientific aims

Complex real-world planning problems involve many (human or automatic) agents interacting cooperatively and robustly via physical communication, sensing actions and dialogue in order to attain common goals in a partially-unpredictable environment. Important aspects of the resulting plan is that it should take into account the beliefs of each agent which can change over time, that it should allow the simultaneous execution of actions, that it should be sufficiently flexible to allow individual agents to make certain choices themselves, and that it should be robust to the failure of certain actions or changes in the environment. In contrast, in classical planning it is supposed that the initial state is completely known, that actions are instantaneous and deterministic and that there are no other agents interfering with the planning agents. On the one hand, one would like to go beyond these restrictive assumptions in order to obtain the required level of expressive power, but on the other hand one would also like to have efficient algorithms solving real-world problems.

The focus of the thesis is the connection between multi-agent planning, explanation and belief revision. Specifically, it focuses on artificial agents interacting with other agents through dialogue. In order to properly achieve their functionalities such agents must be endowed with both planning, explanatory and persuasive capabilities. This includes the agent's capacity (i) to plan a sequence of speech acts aimed at persuading another agent to believe something, (ii) to reply to the other agents questions with appropriate and convincing explanations, (iii) to revise its beliefs during a dialogue.

The novelty of this thesis project lies in a model of explanation in multi-agent planning which can be leveraged for solving real-world problems.

We are interested in applications in which several agents cooperate over a certain duration towards common goals and where all four aspects (multi-agent, temporal, epistemic and contingent) appear together. An example of real-world problem which has these features is planning in the context of human-machine interaction in which an artificial agent is expected to interact with a human user in an informative and persuasive way. To meet these expectations, the agent must be capable of planning a sequence of speech acts aimed at answering to the user's questions and, more generally, at offering to the user a proper understanding of a given domain or situation. Another example is planning in the field of multimode transport. In this context Artificial Intelligence (AI) should be used to propose different combinations of means of transport intelligently in order to offer more economical and greener mobility, closer to the needs of users. This must be achieved according to the availability of different means of transport, current traffic and weather conditions, along with the preferences of the user. Intelligent mobility requires multi-agent coordination (between vehicles, autonomous vehicles, humans, robots, etc.), temporal reasoning (to take into account the durative nature of journeys, actions that may overlap and problems requiring concurrency and synchronization), and epistemic reasoning (to model the dynamic and interactive discovery of knowledge via communication actions to update agents' knowledge). Moreover, an autonomous vehicle must be capable of providing explanations to the user for its choices and behaviors.

This thesis will be held in Artificial Intelligence Department at IRIT, but the PhD student will also benefit from interactions within the MAFTEC[1] working group of the GDR IA whose focus is on multi-agent, flexible, temporal epistemic and contingent planning. It will be supervised with strong interactions with E. Lorini (CNRS Senior Researcher).

2 PhD Workplan

The work to be achieved in the context of the proposed PhD thesis will include original algorithms and/or transformations to simpler versions of planning which have already been widely studied. This may involve the choice of encodings into well-studied problems for which efficient solvers already exist, such as SAT, QBF or SMT (using our tool for problem modelling in logic: TouIST[2] [2]), as well as the redefinition of important notions in classical planning such as landmark actions. This objective will be tackled in different stages of increasing difficulty, leading to the following milestones:

a) Algorithms and software for explanation in multi-agent epistemic planning;
b) Addition of the flexible plans and temporal aspects to these algorithms and software;

[1] https://www.irit.fr/~Frederic.Maris/maftec
[2] https://www.irit.fr/touist

c) Addition of the contingent aspect to these algorithms and software.

3 Expected Results

Among the many examples of application areas that are potentially impacted by this thesis project we can cite: (a) autonomous exploration in possibly hostile environments requiring cooperation between vehicles, humans and robots, (b) service robots including automatic assistance to disabled people, (c) disaster management including emergency evacuation, (d) intelligent mobility, i.e. the use of AI in the field of multi-mode transport to combine and propose more economical and greener mobility.

Our work will lay the foundations for more autonomous agents able to handle more general missions than with current approaches, namely missions involving information gathering, actions upon the environment, partial observability, and coordination with other agents. In a human-centered setting, plans should also cope with privacy concerns which can be expressed by epistemic goals such as "agent 1 must not know the secret of agent 2". Such foundations will be particularly useful for applications like personal assistants and connected objects, especially in domains like health-care, where gathering information (about the patient's health) and respecting privacy are of course primary concerns.

These latter aspects are currently studied in the context of the ANR project CoPains (Cognitive Planning in Persuasive Multimodal Communication).[3] We expect the PhD student to work in connection with the project by elucidating the relationship between planning, explanation and belief revision in human-machine interaction.

4 Required education and background

A master in computer science or equivalent diploma is necessary.
A good knowledge and interests for one or several of the following domains are required: Artificial Intelligence, Combinatorial Algorithmics, Propositional Logic, Complexity Theory.

The obtention of a grant is compulsory and relies on the excellence of the canditate. Full CV (including scholarship results) and motivation letter must be sent to Dominique.Longin@irit.fr, Emiliano.Lorini@irit.fr and Frederic.Maris@irit.fr.

[3] https://www.irit.fr/CoPains/

References

[1] Jorge Luis Fernandez Davila, Dominique Longin, Emiliano Lorini, and Frédéric Maris. A simple framework for cognitive planning. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, A Virtual Conference, February 2-9, 2021*, 2021.

[2] Jorge Fernandez, Olivier Gasquet, Andreas Herzig, Dominique Longin, Emiliano Lorini, Frédéric Maris, and Pierre Régnier. Touist: a friendly language for propositional logic and more. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 5240–5242. ijcai.org, 2020.

[3] Dominique Longin, Emiliano Lorini, and Frédéric Maris. Beliefs, time and space: A language for the yōkai board game. In Takahiro Uchiya, Quan Bai, and Ivan Marsá-Maestre, editors, *PRIMA 2020: Principles and Practice of Multi-Agent Systems - 23rd International Conference, Nagoya, Japan, November 18-20, 2020, Proceedings*, volume 12568 of *Lecture Notes in Computer Science*, pages 386–393. Springer, 2020.

[4] Emiliano Lorini and Francois Schwarzentruber. Multi-agent belief base revision. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 1959–1965. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Main Track.

[5] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019.

[6] Douglas Walton. A new dialectical theory of explanation. *Philosophical Explorations*, 7(1):71–89, 2004.

Dominique LONGIN
CNRS researcher
co-supervisor

Frédéric MARIS
UT3 Associate professor
co-supervisor