Trustworhty AI and Robotics: a Knowledge Engineering perspective

Andrea Orlandini[1], Elisa Foderaro[1], Alessandro Umbrico[1] and Amedeo Cesta[1]

[1]Via S. Martino della Battaglia 44, 00185, Rome. Consiglio Nazionale delle Ricerche, CNR-ISTC, Italy

Abstract

This paper provides an overview of some research activities developed by the Planning and Scheduling Technology (PST) at the Institute of Cognitive Science and Technologies at CNR (CNR-ISTC) addressing the challenges about trust and reliability of AI applications. Specifically, we discuss such issues related in human-robot collaboration in manufacturing scenarios where AI solutions have been deployed.

Keywords

Knowledge Engineering, Trustworthy Artificial Intelligence, Human-Robot Collaboration

1. Introduction

The Planning and Scheduling Technology (PST) Laboratory at Institute of Cognitive Science and Technology was founded in 1997 as a research group focused on Artificial Intelligence (AI) for automatic problem solving with Planning and Scheduling (P&S). The group gathered important results in constraint reasoning, (meta) heuristics for scheduling, mixed-initiative problem solving, timeline-based planning. Since more than twenty years, the group is collaborating with the European Space Agency providing scientific counseling services and developing software technologies to foster the development of autonomous systems and supporting decision making activities. The group developed several research paths working in different areas in National and International projects accumulating a remarkable expertise in AI solutions for assistive robotics, manufacturing and cultural heritage. Many activities concern also technology transfer entailing the need to address challenges related to security, safety, robustness and, more in general, trustworthy AI applications.

More recently, trustworthy AI was defined by the High-level expert group on artificial intelligence[1] appointed by the European Commission to create a set of guidelines for designing, implementing and evaluating AI systems [1]. More specifically, a set of foundational principles and key requirements for trustworthy autonomous systems have been defined considering different issues related to variety of issues like, e.g., ethics, safety, reliability, dependability, security, accurateness, explainability, accountability, etc.

Here, we aim to raise the attention of the reader on a set of orthogonal issues that are under investigation by the authors while developing AI applications. Indeed, in addition to foundational design principles, some subtle barriers usually slower the diffusion of AI-based

---

AIxIA Workshop on Artificial Ingelligence and Robotics (AIRO)

[1] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

solutions in a more practical way. Complex systems often entails the representation of a wide amount of knowledge in AI model. Modeling is always a tricky point and knowledge engineering cannot be handled by non experts. This leads to long and not-so-smooth interactions among different experts to elicit a proper AI model. Also, experts are usually coming in the process with different backgrounds and, often, different perspectives. In this sense, the need of unified procedures is crucial to reduce the risk of misunderstanding and speed up the modeling step. Finally, deploying AI solutions in dynamic conditions is a very critical step and testing and validating solutions is always a nightmare. In this regard, we believe that some specific solutions are needed to facilitate the development and deployment of trustworthy AI solutions such as: i) simplifying AI models definition, ii) creating seamless integration and deployment methodologies for embedded AI technologies and iii) fostering the development of robust deployment. These three aspects are really important for actual implementation of AI systems and strongly related to the elicited foundations and requirements defined by the high-level expert group and deserve to be properly addressed. In this paper, we provide a brief overview of some works performed to address the above issues considering a specific operational scenario, i.e., Human-Robot Collaboration in manufacturing. In particular, we refer to a real-world application in which AI and Robotics technologies are integrated to generate autonomous robotics solutions in human environments such as industries. More in general, the presented results can be applied in different scenarios like, e.g., healthcare assistance and social robotics.

2. AI for Human-Robot Collaboration

Deploying interactive robots in human-populated scenarios requires to address multiple challenges. Among others, it is of paramount importance the ability of the robot to quickly adapt its behaviours to the actual state of the environment to keep the user engaged in the interaction. Such highly flexible and adaptive behaviours are necessary also to guarantee safe and effective human-robot interactions. There are several approaches that aim to achieve robust action selection via planning, e.g., [2, 2, 3, 4] or robust execution via some form of finite state machine (FSM), e.g., [5, 6]. Specifically, the deployment of automated planning techniques brings several advantages in the correctness of the solutions, compactness of the representation, less effort for a designer of the system and more in general, the success of the overall application.

The introduction of tools to facilitate the integration of planning and robotics is then an important step toward advancing the use of AI systems. Nevertheless, the design of well-integrated planning and robotics solutions entails different kind of expertise to address a wide variety of control issues, spanning from low-level control to decisional autonomy configuration. The design of plan-based autonomous robots entails domain, robotics and planning experts interacting and sharing different kind of knowledge and techniques, often pursuing different control perspectives, and tightly collaborating to define integrated models compelling a wide set of requirements. Apart technological limitations, a set of knowledge engineering problems can be clearly identified: information sharing at different abstraction levels may cause redundancies or even inconsistencies in planning specifications; the lack of a generally accepted design methodology may entail many potential back-and-forth (re)work over models and control parameters before defining the proper control configuration; state-of-the-art software tools are usually developed to support either robotics or planning experts and not the overall process as a collaborative work.

2.1. Modeling

The development of tools that facilitate the integration (and interaction) of AI planning and robotics entails different skills that are all necessary to effectively address the underlying variety of control issues, spanning from low-level control to decisional (and behavioral) autonomy [7]. Among others, a crucial knowledge engineering problem is the lack of a generally accepted modeling methodology entailing many potential back-and-forth (re)work over models and control parameters before defining the proper control configuration.

Some attempts to connect AI and Robotics environments have been made. For instance, ROSPlan [2] has been proposed as a unique integrated solution for PDDL-based planners to be smoothly deployed in ROS architectures [8] ROSPlan constitutes a well-integrated solution and has been used in various robotic domains for this purpose [8, 9, 10] also with planning techniques specifically used for human-robot interaction [11]. Robotics experts can easily connect their ROS-based modules to any (supported) PDDL-planner but there is no support to define planning specifications. So, roboticists are left alone in managing a knowledge planning modelling. Several timeline-based planning frameworks such as, e.g., EUROPA [12] and APSI-KEEN [13], provide knowledge engineering support for planning. But none of them provides structured support for deployment of robotic applications. In general, all those solutions require robotic experts to have some expertise in planning specification.

Some solutions addressing this issue have been proposed to facilitate the interaction between AI and Robotics experts. Yet, a domain expert may have difficulties in approaching this kind of solutions. A *Domain expert* is usually responsible for the definition of the tasks and overall goals of a robotic system, while other actors have responsibility on more specific aspects, i.e., a *planning expert*, with models to provide robust A.I. planning features, and a *robotic expert*, responsible for implementing robot operations.

A software tool, called "Tool fostEriNg Ai plaNning in roboTics" (TENANT) [14], is addressing the needs of Domain Experts to set goals, defining tasks and set operational constraints notwithstanding the intrinsic complexity required at planning and robotics level. TENANT is a general purpose tool that can be deployed for addressing multiple applications/domains and can be easily integrated with Planning and Scheduling software framework, e.g., PLATINUm [15]. TENANT specifically focuses on Human-Robot Collaboration and allows domain experts to specify *collaborative models* and thus describe *tasks* needed to achieve desired (productive) goals. Tasks can be either compound or simple (i.e., further structured in other sub-tasks) and are characterized by specific configurations/properties relevant for their execution, like e.g., collaborative modalities, or assignment preferences specifying who is actually in charge of performing a certain (sub)task. Furthermore, TENANT supports the definition of *operational constraints* such as, e.g., temporal synchronizations or precedence constraints in order to characterize the correct execution of the resulting production/control flow. Given a complete (and correct) collaborative model, TENANT can automatically generate a suitable *planning model* that can be used to feed a Planning & Scheduling system thus enabling intelligent (collaborative) robot behaviours.

TENANT was validated on a concrete (and realistic) HRC production process derived from an EU-funded research project called Sharework[2]. The validation shows the feasibility of the tool and the underlying engineering methodology.

2.2. Integration and Deployment

In order to foster integration and deployment, two main software frameworks have been developed: ROXANNE and ROS-TiPlEx.

ROXANNE (ROs fleXible ActiNg coNtrollEr) is an FTP ROSIN project[3] developing ROS packages that facilitate the integration of Artificial Intelligence planning and execution capabilities with robotic platforms. ROXANNE specifically supports the development of timeline-based goal-oriented architectures in ROS. The project aim at integrating of timeline-based planning and execution technologies with ”standard” robot control techniques to enhance robustness and flexibility of robot behaviors when dealing with uncontrollable dynamics of an environment or other ”external” and concurrent agents like e.g., human operators in Human-Robot Collaboration (HRC) scenarios.

ROXANNE aims at creating a ROS compliant framework facilitating the use of timeline-based planning and execution capabilities in industrial settings. Target users are either researchers and manufacturing companies that want to evaluate flexible task-level controllers to better support production processes. The objectives of the project are thus the following:

- Facilitate industrial robot programming by providing a general-purpose specification language to model operational and temporal constraints of production processes and the dynamics of working environment, human operators and robot behaviors.
- Realize a ROS-integrated goal-oriented planning and execution system capable of autonomously synthesize the set of tasks a robot must perform to achieve production goals. Minimize production interruptions by dynamically coordinating and adapting robot behaviors by taking into account uncontrollable dynamics of a working environment.
- Provide process-driven modeling and robot-agnostic control capabilities to support interoperability and integration with different robotic platforms.
- Define an interoperability communication protocol characterizing events and information exchanged within the life-cycle of a dynamic task planning system. Such a protocol defines services and dependencies that are necessary for the effective integration of the task planner with robot controllers and realize an autonomous robot architecture.
- Enable and facilitate the use of timeline-based control techniques in real-world production contexts by leveraging a standard platform like ROS and an existing timeline-based framework called PLATINUm.

ROS-TiPlEx. A novel comprehensive framework, called ROS-TiPlEx (*Timeline-based Planning and Execution with ROS*) [7], is to provide a shared environment in which experts in robotics

---

[2] http://wwww.sharework-project.eu

[3] ROXANNE was funded by the European Union’s Horizon 2020 research and innovation programme under the project ROSIN with the gran agreement No 732287.

and planning can easily interact to, respectively, encode information about low-level robot control and define task planning and execution models. The long-term goal of ROS-TiPlEx is to provide a standardized modelling process in the form of a tool; such a tool will support experts during the overall robot control configuration and will shape the process to ensure quality and safety, and also will improve the productivity. Namely, ROS-TiPlEx is to support the abstraction steps facilitating the interaction between domain experts and engineers in two independent fields (ROS and Automated Planning) to provide a complete control software configuration. ROS-TiPlEx is also to store all the needed information and quickly adapt robot and software configuration to different contexts.

To endow autonomous robot systems with the ability to perform dynamic task planning, control architectures can be equipped with planning software. This entails the definition of suitable task planning models to capture the significant elements of cooperative missions (e.g., exploration tasks) and the configuration of planning software for robust task plan execution. The process can be described as follows: all the robotic components configuration must be modelled as an abstraction for task planning (e.g., timed automata and temporal constraints). Then, an automated planning software, receiving as input such specification, can guarantee the achievement of general mission objectives (defined by a domain expert) producing as output a sequence of planned tasks that, if dispatched back to the robot control software and correctly executed, will ensure the achievement of the mission goals. Such a process requires a non-trivial effort to connect two very different and independent modules, i.e., a robot control module and an automated planning module as well as offer an easy to use interface to domain experts for defining general mission goals.

ROS-TiPlEx aims at providing a standardized modelling process supporting these experts during the overall robot configuration phase and guiding the overall process to foster mission quality and safety as well as improve its productivity. In fact, in an HRI scenario, an effective planning domain must take into account many features: (i) guarantee human safety, (ii) increase the effectiveness of robot tasks, (iii) maximize (as much as possible) the utility (according to a given function) of the whole system. It is clear that to achieve the above objectives, a robust planning domain is needed. In this regard, ROS-TiPlEx supports the abstraction steps to facilitate the interaction between the two independent expertise fields (ROS and automated planning), to provide a complete control software configuration. While an interface for domain experts is not yet included, another asset of ROS-TiPlEx is the ability to store useful information about the modelling in a local DB, to promote reusability and facilitate any mission requirement change. In this paper, we focus on interactions among planning and robotics experts considering the mission objectives pre-defined by the domain expert as given. In the future, we will extend the tool to consider also domain experts as additional actors in the process.

The proposed ROS-TiPlEx workflow begins with a *robot expert* who provides the design and the robot configuration in ROS, being thus able to control the robot elements via, e.g., robot programming. The *robot expert* is in charge of mapping the very basic capabilities of the robotic platform, sensors, actuators and payloads to the first layer of abstraction in ROS-TiPlEx. This process generates as output a set of atomic actions aligned with the mission objectives, that will be all stored in a local DB. Consequently, a *planning expert* can access the information in the local DB and then provide a further abstraction based on those data and is in charge of mapping the relevant elements of the first abstraction layer such as, e.g., sensors readings, set of robot atomic actions, etc., in a planning specification. The result of the rework is a suitable robot planning model. ROS-TiPlEx proposes a defined general process implemented as a tool with accessible graphical interfaces (one for robot experts and one for planning experts), separating the underlying logic from the configuration process. Such a tool is to offer a single access point, presenting to each kind of expert a role-specific vision on the overall system, allowing them to manage the data closer to their expertise. In this way, the two kind of experts do not need to build strong cross-competencies or to have long iterative interaction to build a shared model, because the fixed structure of information and processes provided by ROS-TiPlEx prevents possible misunderstanding and secures the process.

2.3. Dynamic Operation and Execution

In such real world contexts, Artificial Intelligence techniques can bring effective solutions to foster autonomy and effective control. For example, dynamic activity planning systems based on flexible solutions can be a Key Enabling Technology for HRC controllers in which the movements of the robots must be continuously adapted to the presence of human beings acting as uncontrollable ”agents” in the environment. Their presence requires the ability for control systems to evaluate the variability of robot execution times and, in this sense, standard control methods are not entirely effective. Furthermore, the integration of Planning and Scheduling (P&S) technology with Knowledge Engineering solutions and, more specifically, with Verification and Validation (V&V) techniques is a key element to synthesize safety-critical systems in robotics [13, 16]. For over a decade, the ISTC-CNR PST group has launched a research initiative to study the possible integration of a timeline-based planning framework (APSI [17]) and V&V techniques based on Timed Game Automata (TGA) to automatically synthesize a robot controller that guarantees robustness and safety properties [18, 19]. Indeed, some control systems are based on time planning mechanisms capable of managing coordinated activities with a certain temporal flexibility (uncertainty) (for example, [20]) that exploit time planners (for example , [12]). Unfortunately, these systems do not allow for an explicit representation of uncontrollable characteristics. Consequently, the controllers developed with these technologies do not have the *robustness* necessary to cope with the temporal uncertainty of HRC scenarios and the resulting [21] controllability problems. These systems are usually based on reprogramming mechanisms which can however severely penalize production performance. The goal of this long-term research is to build a robust activity planning system that allows for a flexible, safe and efficient HRC. In [22, 16], a general approach pursued in some research projects was presented with the aim of creating controllers for collaborative robots capable of dynamically coordinating production tasks based on the behavior of human workers.

3. Conclusions

This paper provided an overview of the work done by the PST research group at the Institute of Cognitive Science and Technologies at CNR (CNR-ISTC). Several results and software tools were presented with the aim of addressing trust and reliability issues in AI and Robotics applications. A knowledge engineering perspective was pursued to emphasize some aspects usually less considered.

Acknowledgments

Authors are partially supported by SHAREWORK (H2020 Factories of the Future GA No. 820807) and TAILOR (H2020 Towards a vibrant European network of AI excellence centres, GA 952215).

References

[1] High-Level Expert Group on AI, Ethics guidelines for trustworthy AI, Report, European Commission, Brussels, 2019. URL: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[2] M. Cashmore, M. Fox, D. Long, D. Magazzeni, B. Ridder, A. Carrera, N. Palomeras, N. Hurtos, M. Carreras, Rosplan: Planning in the robot operating system, in: Proceedings of the International Conference on Automated Planning and Scheduling, volume 25, 2015.

[3] C. P. C. Chanel, C. Lesire, F. Teichteil-Königsbuch, A robotic execution framework for online probabilistic (re)planning, in: Proceedings of ICAPS, AAAI press, 2014.

[4] B. Lacerda, D. Parker, N. Hawes, Optimal and dynamic planning for markov decision processes with co-safe ltl specifications, in: IEEE International Conference on Intelligent Robots and Systems, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 1511–1516. doi:10.1109/IROS.2014.6942756.

[5] J. Bohren, S. Cousins, The smach high-level executive [ros news], IEEE Robotics Automation Magazine 17 (2010) 18–20. doi:10.1109/MRA.2010.938836.

[6] V. A. Ziparo, L. Iocchi, P. U. Lima, D. Nardi, P. F. Palamara, Petri net plans, Autonomous Agents and Multi-Agent Systems 23 (2011) 344–383.

[7] C. La Viola, A. Orlandini, A. Umbrico, A. Cesta, Ros-tiplex: How to make experts in a.i. planning and robotics talk together and be happy, in: 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 2019.

[8] M. Cashmore, M. Fox, T. Larkworthy, D. Long, D. Magazzeni, Auv mission control via temporal planning, in: Proceedings - IEEE International Conference on Robotics and Automation, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 6535–6541. doi:10.1109/ICRA.2014.6907823.

[9] R. I. Brafman, M. Bar-Sinai, M. Ashkenazi, Performance level profiles: A formal language for describing the expected performance of functional modules, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, pp. 1751–1756. doi:10.1109/IROS.2016.7759280.

[10] M. Cashmore, M. Fox, D. Long, D. Magazzeni, B. Ridder, Opportunistic planning in autonomous underwater missions, IEEE Transactions on Automation Science and Engineering 15 (2018) 519–530. doi:10.1109/TASE.2016.2636662.

[11] V. Sanelli, M. Cashmore, D. Magazzeni, L. Iocchi, Short-term human-robot interaction through conditional planning and execution, in: Proceedings of ICAPS, AAAI press, 2017.

[12] J. Barreiro, M. Boyce, M. Do, J. Frank, M. Iatauro, T. Kichkaylo, P. Morris, J. Ong, E. Remolina, T. Smith, D. Smith, EUROPA: A Platform for AI Planning, Scheduling, Constraint Programming, and Optimization, in: ICKEPS 2012: the 4th Int. Competition on Knowledge Engineering for Planning and Scheduling, 2012.

[13] A. Orlandini, G. Bernardi, A. Cesta, A. Finzi, Planning meets verification and validation in a knowledge engineering environment, Intelligenza Artificiale 8 (2014) 87–100.

[14] E. Foderaro, A. Cesta, A. Umbrico, A. Orlandini, Simplifying the a.i. planning modeling for human-robot collaboration, in: 2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN), 2021, pp. 1011–1016. doi:`10.1109/RO-MAN50785.2021.9515431`.

[15] A. Umbrico, A. Cesta, M. Cialdea Mayer, A. Orlandini, Platinum: A new framework for planning and acting, in: AI*IA 2017 Advances in Artificial Intelligence, 2017, pp. 498–512.

[16] A. Orlandini, M. Cialdea Mayer, A. Umbrico, A. Cesta, Design of Timeline-Based Planning Systems for Safe Human-Robot Collaboration, Springer International Publishing, Cham, 2020, pp. 231–248. URL: https://doi.org/10.1007/978-3-030-38561-3_12. doi:`10.1007/978-3-030-38561-3_12`.

[17] A. Cesta, G. Cortellessa, S. Fratini, A. Oddi, MrSPOCK: Steps in Developing an End-to-End Space Application, Computational Intelligence 27 (2011).

[18] A. Cesta, A. Finzi, S. Fratini, A. Orlandini, E. Tronci, Flexible Timeline-Based Plan Verification, in: KI 2009: Advances in Artificial Intelligence, volume 5803 of *LNAI*, 2009.

[19] A. Orlandini, M. Suriano, A. Cesta, A. Finzi, Controller synthesis for safety critical planning, in: IEEE 25th International Conference on Tools with Artificial Intelligence (ICTAI 2013), IEEE, 2013, pp. 306–313.

[20] F. Py, K. Rajan, C. McGann, A systematic agent framework for situated autonomous systems, in: AAMAS, 2010, pp. 583–590.

[21] P. H. Morris, N. Muscettola, Temporal Dynamic Controllability Revisited, in: Proc. of AAAI 2005, 2005, pp. 1193–1198.

[22] A. Cesta, A. Orlandini, G. Bernardi, A. Umbrico, Towards a planning-based framework for symbiotic human-robot collaboration, in: 21th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, 2016.