In *Proceedings of the ICAPS'17 Workshop on Planning and Robotics,*
Pittsburgh, USA June 2017

Integrated Commonsense Reasoning and Probabilistic Planning

Shiqi Zhang[†] **and Peter Stone**[‡]
[†]Department of Electrical Engineering and Computer Science, Cleveland State University
[‡]Department of Computer Science, The University of Texas at Austin
s.zhang9@csuohio.edu; pstone@cs.utexas.edu

Abstract

Commonsense reasoning and *probabilistic planning* are two of the most important research areas in artificial intelligence. This paper focuses on Integrated commonsense Reasoning and probabilistic Planning (**IRP**) problems. On one hand, commonsense reasoning algorithms aim at drawing conclusions using structured knowledge that is typically provided in a declarative way. On the other hand, probabilistic planning algorithms aim at generating an action policy that can be used for action selection under uncertainty. Intuitively, reasoning and planning techniques are good at "understanding the world" and "accomplishing the task" respectively. This paper discusses the complementary features of the two computing paradigms, presents the (potential) advantages of their integration, and summarizes existing research on this topic.

Introduction

Robots that operate in the real world frequently need to work on complex tasks that require more than one action. Two planning paradigms have been developed for robots that work on such complex tasks: *task planning* and *probabilistic planning*. Task planning algorithms focus on computing a *sequence* of actions, implicitly assuming perfect action executions in a deterministic domain. Probabilistic planning algorithms aim at, in stochastic domains, computing an action *policy* that suggests an action from any state under the uncertainty from the non-deterministic outcomes of robot actions. Examples of non-deterministic action outcomes include opponent moves in chess and results of grasping an object using an unreliable gripper. This paper focuses on *probabilistic planning* in stochastic domains.

The Markov assumption states that the next state only relies on the current state and is independent of all previous states (the first-order case). Accordingly, Markov decision processes (MDPs) and partially observable MDPs (POMDPs) have been developed as probabilistic planning frameworks under full and partial observabilities respectively (Kaelbling, Littman, and Cassandra 1998). When the current world state is not directly observable, the robot needs to make observations to estimate the current state, where the observations are frequently local and unreliable. Accordingly, a belief distribution over all possible states is maintained as the state estimation representation. MDP and POMDP algorithms, e.g., value iteration (Sutton and Barto 1998), Monte Carlo tree search (Kocsis and Szepesvári 2006) and SARSOP (Kurniawati, Hsu, and Lee 2008), help compute a *policy* that enables planning toward maximizing long-term rewards.

Orthogonal to planning, commonsense knowledge is used to refer to the knowledge that is normally true but not always. Such knowledge can be represented in different forms, e.g., as defaults and using probabilities. Commonsense reasoning is concerned with drawing conclusions (or generating new knowledge) using the existing commonsense knowledge. Generally speaking, all knowledge is commonsense knowledge and can be represented in very different forms, such as First-Order Logic (FOL) (Smullyan 1995), Markov Logic Networks (MLNs) (Richardson and Domingos 2006), and Answer Set Programming (Gelfond and Kahl 2014). Such reasoning paradigms are good at drawing (deterministic, probabilistic, or both) conclusions within a static world, but is ill-equipped for planning to achieve long-term goals in dynamic, stochastic domains.

The difficulty of solving MDP and POMDP problems comes from the two major computational challenges of "curse of dimensionality" (a complex robotic task generates a high-dimensional state space) and "curse of history" (a robot often needs to take many actions to reach the goal, resulting in a long planning horizon) (Kurniawati et al. 2011).

> *The main objective of Integrated commonsense Reasoning and probabilistic Planning (IRP) algorithms is to decompose a robot planning problem into two sub-problems: commonsense reasoning and probabilistic planning. Then a commonsense reasoner and a probabilistic planner can be used to focus on the sub-problems of high-dimensional reasoning and long-horizon planning respectively. .*

In what follows, we first present a state space decomposition strategy that paves the way of IRP methods, and then summarize existing research related to this topic.

State Space Decomposition

State space decomposition plays an important role in IRP algorithms. We first define **endogenous** and **exogenous** domain variables for the sake of easier discussion. Endogenous variables are the variables whose values the robot wants to *actively* change or observe (or both). Exogenous variables

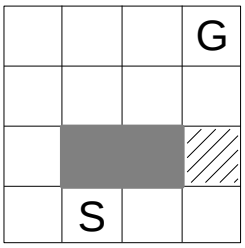

Figure 1: An illustrative example: the robot needs to navigate from its start location (*S*) to the goal (*G*). The hatching area on the right is a near-window area where the robot can be trapped (probabilistically) under sunlight.

are the variables whose values the robot only wants to *passively* observe and adapt to as needed.

Consider a robot navigation problem in a fully-observable 2D grid world shown in Figure 1. The robot can take actions (*North*, *East*, *South*, and *West*) to move toward one of its nearby grid cells, and such actions succeed probabilistically. The hatching cell is a dangerous area to the robot, because, in the mornings, sunlight there can blind its range-finder sensor, causing it unrecoverably lost (probabilistically). In this example, the robot's current location should be modeled as an *endogenous* variable, because its value change needs to be modeled in the planning process, i.e., its value needs to be *actively* changed. Current time (morning or not) should be modeled as an *exogenous* variable, meaning that the robot does not need to change its value in the planning process. However, it is indeed necessary to keep an eye on (*passively* observe) its value, and adjust the probabilistic planner as needed, e.g., reducing the success rate of navigating though the near-window cell when current time is morning.

In principle, all domain variables should be modeled in (PO)MDPs. However, in practice, we usually do not do that, because there is always the trade-off between model completeness and computational tractability. The goal of maintaining two sets of variables is to enable the robot to focus on planning over a long horizon in a relatively small state space (partial space) and reasoning within a relatively large state space (full space). Given full and partial state spaces where the robot reasons and plans respectively, the question will be how the reasoning and planning in two different spaces are connected, which will be discussed next.

Existing Research on IRP Problems

Logical commonsense reasoning has been incorporated into probabilistic planning to compute an informative prior (Zhang, Sridharan, and Bao 2012; Zhang, Sridharan, and Wyatt 2015). In that work, a target search problem was used as the application domain. The robot's noisy observations were modeled using a POMDP, and the belief distribution of the POMDP represents the estimate of the target's position, as the single endogenous variable. The robot moves to different areas in a large office domain to "uncover" the position of the target object. A categorical tree that includes a large number of exogenous variables (such as scanners and printers are office electronics) was constructed using a logical reasoner. As a result, the probabilistic planner is able to focus on a very small partial space that includes only the variable of the target's position, while being able to reason about the target's likely positions within a much larger state space. The gap between commonsense reasoning and probabilistic planning was bridged by using a set of heuristics (such as printers are usually collocated with scanners) to convert deterministic conclusions into a distribution for a POMDP.

In order to better bridge the gap between commonsense reasoning and probabilistic planning, some IRP algorithms have used reasoners that are able to reason about both logical and probabilistic commonsense knowledge. These algorithms and implementations include CORPP (Zhang and Stone 2015) and OpenDial (Lison 2015) that use P-log (Baral, Gelfond, and Rushton 2009) and MLN (Richardson and Domingos 2006) for commonsense reasoning respectively. Their commonsense reasoners are able to directly output a probability distribution for the planner. For instance, a spoken dialog problem was used as the application domain in (Zhang and Stone 2015), where the robot uses unreliable speech recognition to identify the human's request. In that work, the state space decomposition enables the probabilistic planner to focus on only the endogenous variables that are needed for specifying the requests (such as delivering *coffee* for *alice*). All other variables, such as time – people prefer buying coffee in the *mornings*, are modeled as exogenous variables and handled by the commonsense reasoner.

There are other ways of integrating commonsense reasoning and probabilistic planning, where full and partial state spaces are not explicitly differentiated. A refinement-based architecture has been developed for robot reasoning and planning (Sridharan et al. 2015). At the high level, an action language is used for computing a sequence of symbolic actions to deterministically guide the robot behaviors. At the low level, a probabilistic model (a POMDP) is used for physically implementing these actions. As a result, in that work, the high level reasoning layer is able to conduct complicated reasoning tasks, such as explaining history behaviors, that are impossible for probabilistic planners. In another line of research, commonsense reasoning was used for diagnostic tasks and generating explanations, and a hybrid planner allows switching between deterministic and probabilistic planners (Hanheide et al. 2015). POMDP-based planning has been integrated with commonsense learning, where the agent learns from a set of example traces and commonsense knowledge refers to the knowledge based on which a reference policy generates the example traces (Juba 2016).

Probabilistic planning frameworks and algorithms assume a known world model (including world dynamics and robot capabilities). In case of an unknown world, reinforcement learning (RL) algorithms can be used to help an agent learn an action policy by interacting with the environments (Sutton and Barto 1998). Existing research has studied the integration of commonsense reasoning and RL. For instance, relational RL has been used for learning robot action precon-

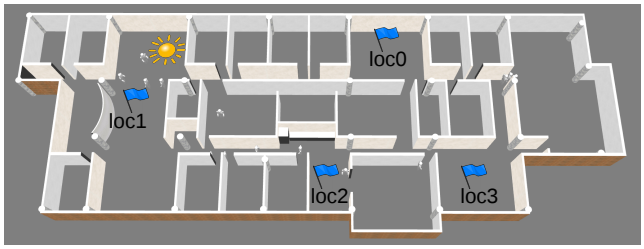

Figure 2: The robot navigation domain that includes four possible navigation goals. Human pedestrians might block the hallway (probabilistically), and sunlight can blind the robot's range-finder sensors (probabilistically).

ditions (affordances), as a kind of commonsense knowledge about robot capabilities (Sridharan, Meadows, and Gomez 2017). In order to reduce the space of exploration in RL, a commonsense reasoner has been used to help the agent to focus on only the *reasonable* actions, significantly increasing the learning rate (Leonetti, Iocchi, and Stone 2016).

In what follows, we summarize our IRP algorithm called iCORPP that dynamically constructs (PO)MDPs to shield exogenous variables from (PO)MDPs while still enabling probabilistic planning to adapt to the exogenous events.

A Summary of iCORPP, an IRP Algorithm

A general description of interleaved CORPP (iCORPP) is available in our recent paper (Zhang, Khandelwal, and Stone 2017). In this section, we directly present an instantiation of iCORPP on a robot navigation problem, and compare it against CORPP, which is similar except that CORPP requires the planner to consider any exogenous variables that could change its transition dynamics. Figure 2 shows the domain map, where the robot needs to visit the four locations that are connected through a corridor. However, human pedestrians can block the way (probabilistically) in the corridor and sunlight can blind the robot's range-finder sensors. It is also known that sunlight only exists in near-window areas when the time is morning and the weather is sunny.

We assume the values of all domain variables are fully observable, so we can use an MDP to construct the planner. If we model only thirty locations in the corridor, there will be ten states in the state space. When we consider each of the locations can be either occupied or unoccupied by humans, the number of states becomes 30×2^{30}. When we further consider each of the locations can be either under sunlight or not, the number of states becomes $30 \times 2^{30} \times 2^{30}$, which is a huge number. This is a small toy domain, and we still have not considered the domain variables of time, weather, and each area is near-window or not.

The whole idea of iCORPP in this domain is to *model only robot position as the endogenous variable for probabilistic planning and all others as exogenous variables to be considered only by the commonsense reasoner.*

Next, we very briefly describe our commonsense reasoner, where the probabilistic transition system of MDP is described in P-log (Baral, Gelfond, and Rushton 2009). In case of exogenous events, our commonsense reasoner dynamically constructs a new MDP that captures the effects of the exogenous variables on the transition dynamics of the endogenous variables.

The navigation domain shown in Figure 2 is defined using sorts `row` and `col`, and predicates `belowof` and `leftof`. We then introduce predicates `near_row` and `near_col` used for specifying if two grid cells are next to each other, where `R`'s (`C`'s) are variables of row (column).

$$\texttt{near\_row}(\texttt{RW}_1,\texttt{RW}_2) \leftarrow \texttt{belowof}(\texttt{RW}_1,\texttt{RW}_2).$$
$$\texttt{near\_row}(\texttt{RW}_1,\texttt{RW}_2) \leftarrow \texttt{near\_row}(\texttt{RW}_2,\texttt{RW}_1).$$
$$\texttt{near\_col}(\texttt{CL}_1,\texttt{CL}_2) \leftarrow \texttt{leftof}(\texttt{CL}_1,\texttt{CL}_2).$$
$$\texttt{near\_col}(\texttt{CL}_1,\texttt{CL}_2) \leftarrow \texttt{near\_col}(\texttt{CL}_2,\texttt{CL}_1).$$

We use predicates `near_window` and `sunny` to define the cells that are near to window and the cells that are actually under sunlight. The rule below is a default stating that: in the mornings, a cell near window is believed to be under sunlight, unless defeated elsewhere.

$$\texttt{sunny}(\texttt{RW},\texttt{CL}) \leftarrow \texttt{near\_window}(\texttt{RW},\texttt{CL}),\ \texttt{not}\ \neg\texttt{sunny}(\texttt{RW},\texttt{CL}),\ \texttt{curr\_time} = \texttt{morning}.$$

While navigating in areas under sunlight, there is a large probability of becoming lost (0.9), which deterministically leads to the end of an episode.

$$\texttt{pr}(\texttt{next\_term} = \texttt{true} \mid \texttt{curr\_row} = \texttt{RW},\ \texttt{curr\_col} = \texttt{CL},\ \texttt{sunny}(\texttt{RW},\texttt{CL}),\ \texttt{curr\_term} = \texttt{false}) = 0.9.$$
$$\texttt{pr}(\texttt{next\_term} = \texttt{true} \mid \texttt{curr\_term} = \texttt{true}) = 1.0.$$

The robot can take actions to move to a grid cell next to its current one: $\texttt{action} = \{\texttt{left},\texttt{right},\texttt{up},\texttt{down}\}$. For instance, given action `up`, the probability of successfully moving to the above grid cell is 0.9, given no obstacle in the above cell.

$$\texttt{pr}(\texttt{next\_row} = \texttt{RW}_2 \mid \texttt{curr\_row} = \texttt{RW}_1,\ \texttt{curr\_col} = \texttt{CL}_1,\ \texttt{belowof}(\texttt{RW}_1,\texttt{RW}_2),\ \neg\texttt{sunny}(\texttt{RW}_2,\texttt{CL}_1),\ \neg\texttt{blocked}(\texttt{RW}_2,\texttt{CL}_1),\ \texttt{curr\_a} = \texttt{up}) = 0.9.$$

iCORPP significantly reduces the complexity of probabilistic planning compared to its one-shot solution, while enabling robot behaviors to adapt to exogenous changes. As an example on complexity, the MDP constructed by iCORPP (thirty positions, five weather conditions and three times) includes only 60 states, whereas the traditional way of enumerating all combinations of attribute values (Boutilier, Dean, and Hanks 1999), produces more than 2^69 states, which cannot be solved (accurately or approximately) in practice.

Experimental Results

Experiments in simulation were conducted using GAZEBO (Koenig and Howard 2004). We used a solver introduced in (Zhu 2012) for P-log programs (except that reasoning about reward was manually conducted) and value iteration for MDPs (Sutton and Barto 1998).

We limit the number of random walkers to be 1 and its speed to be one fifth of the robot's. A goal room is randomly selected from the four flag rooms. Reasoning happens only after the current episode is terminated (goal room is reached). The walker's position is the only exogenous domain change (by temporarily setting the time to be "evening"). We cached policies for both CORPP as the baseline (4 policies) and iCORPP (56 policies).

The walker moves slowly between $loc0$ and $loc2$. Without adaptive planning developed in this work, the robot follows the "optimal" path and keeps trying to bypass the walker for a fixed length of time. If the low-level motion planner does not find a way to bypass the walker within the time, the robot will take the other way to navigate to the other side of the walker and continues executing the "optimal" plan generated by the outdated model. When the robot navigates between $loc0$ and $loc2$, iCORPP reduces the traveling time from about 250 seconds to about 110 seconds, producing a significant improvement.

A comprehensive description of the experimental results is available in our iCORPP paper (Zhang, Khandelwal, and Stone 2017) and this web page includes videos of real-robot experiments. [1]

Conclusions

In this paper, we present the motivation of Integrated commonsense Reasoning and probabilistic Planning (IRP) within the context of robot planning. We summarize existing research on this topic and present our recent work, called iCORPP, that dynamically constructs MDPs and POMDPs for adaptive robot planning. The general idea of IRP algorithms is to decompose the original probabilistic planning problems into the sub-problems of commonsense reasoning and probabilistic planning that respectively focus on "understanding the world" and "accomplishing the task". iCORPP demonstrates that this decomposition significantly reduces the state space where planning is conducted and enables robot to adapt to the value change of exogenous variables without including these variable in planning models.

Acknowledgments

A portion of this work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CNS-1330072, CNS-1305287, IIS-1637736, IIS-1651089), ONR (21C184-01), AFOSR (FA9550-14-1-0087), Raytheon, Toyota, AT&T, and Lockheed Martin. Peter Stone serves on the Board of Directors of, Cogitai, Inc. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

References

Baral, C.; Gelfond, M.; and Rushton, N. 2009. Probabilistic Reasoning with Answer Sets. *Theory and Practice of Logic Programming* 9(1):57–144.

Boutilier, C.; Dean, T.; and Hanks, S. 1999. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research* 11(1):94.

Gelfond, M., and Kahl, Y. 2014. *Knowledge Representation, Reasoning, and the Design of Intelligent Agents: The Answer-Set Programming Approach*. Cambridge University Press.

Hanheide, M.; Göbelbecker, M.; Horn, G. S.; Pronobis, A.; Sjöö, K.; Aydemir, A.; Jensfelt, P.; Gretton, C.; Dearden, R.; Janicek, M.; et al. 2015. Robot task planning and explanation in open and uncertain worlds. *Artificial Intelligence*.

Juba, B. 2016. Integrated common sense learning and planning in pomdps. *Journal of Machine Learning Research* 17(96):1–37.

Kaelbling, L. P.; Littman, M. L.; and Cassandra, A. R. 1998. Planning and acting in partially observable stochastic domains. *Artificial intelligence* 101(1):99–134.

Kocsis, L., and Szepesvári, C. 2006. Bandit based monte-carlo planning. In *Machine Learning: ECML 2006*. Springer. 282–293.

Koenig, N., and Howard, A. 2004. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Kurniawati, H.; Du, Y.; Hsu, D.; and Lee, W. S. 2011. Motion planning under uncertainty for robotic tasks with long time horizons. *The International Journal of Robotics Research* 30(3):308–323.

Kurniawati, H.; Hsu, D.; and Lee, W. S. 2008. Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. In *Robotics: Science and Systems*.

Leonetti, M.; Iocchi, L.; and Stone, P. 2016. A synthesis of automated planning and reinforcement learning for efficient, robust decision-making. *Artificial Intelligence* 241:103–130.

Lison, P. 2015. A hybrid approach to dialogue management based on probabilistic rules. *Computer Speech & Language* 34(1):232–255.

Richardson, M., and Domingos, P. 2006. Markov logic networks. *Machine Learning* 62(1-2):107–136.

Smullyan, R. M. 1995. *First-order logic*. Courier Corporation.

Sridharan, M.; Gelfond, M.; Zhang, S.; and Wyatt, J. 2015. A refinement-based architecture for knowledge representation and reasoning in robotics. *arXiv preprint arXiv:1508.03891*.

Sridharan, M.; Meadows, B.; and Gomez, R. 2017. What can i not do? towards an architecture for reasoning about and learning affordances. In *Proceedings of International Conference on Automated Planning and Scheduling (ICAPS)*.

Sutton, R. S., and Barto, A. G. 1998. *Reinforcement learning: An introduction*. MIT press Cambridge.

Zhang, S., and Stone, P. 2015. CORPP: Commonsense reasoning and probabilistic planning, as applied to dialog with a mobile robot. In *Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, 1394–1400.

Zhang, S.; Khandelwal, P.; and Stone, P. 2017. Dynamically constructed (po)mdps for adaptive robot planning. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*.

Zhang, S.; Sridharan, M.; and Bao, F. S. 2012. ASP+POMDP: Integrating Non-monotonic Logic Programming and Probabilistic Planning on Robots. In *International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob)*.

Zhang, S.; Sridharan, M.; and Wyatt, J. L. 2015. Mixed logical inference and probabilistic planning for robots in unreliable worlds. *IEEE Transactions on Robotics* 31(3):699–713.

Zhu, W. 2012. *PLOG: Its Algorithms and Applications*. Ph.D. Dissertation, Texas Tech University, USA.

[1] http://eecs.csuohio.edu/~szhang/corpp/