Exploration/Exploitation in Path Planning Using Probability Propagation *

Francesco A. N. Palmieri,† Krishna R. Pattipati,‡ Giovanni Fioretti
Francesco Verolla, Giovanni Di Gennaro,§ Amedeo Buonanno¶

Abstract

We apply probability propagation to the determination of paths in discrete grids with agents' dynamics modeled as Markov Decision Processes. The probability flow is used to determine best solutions when there is only partial knowledge of the obstacle map and the goal. Distributed probability is used in the backward flow to attract the agent towards unknown regions in a sequence of exploration/exploitation steps.

1. Introduction

Modeling the behavior of moving human agents is currently of great interest both in surveillance and in robotics. A recent review can be found in [1]. Some of the authors of this paper have also contributed to the topic [2][3] with various proposals. The focus of this paper is on probability-based models, as they seem to capture in a natural way the uncertainties and the constraints imposed by a real scenario.

The original idea to determine paths with a probability model is due to Attias [6]. Since then, probability models have received considerable attention, for example as in [4], [5]. Attias proposed that, for a system with dynamics described by a Markov model, conditioning on initial and final states (start and goal), and performing inference on the intermediate states and actions, can provide a constrained path solution. This method has been also shown to be directly related to stochastic dynamic programming [7] [8]. Some of these ideas are related to free-energy stochastic models [9] [10] [11] and to KL-learning [12] and are currently under active investigation. In our view, the probabilistic paradigm seems quite appealing, as it may be one of the best approaches towards a unified view on modeling intelligent behavior.

In [13], using probability message propagation, we have simulated single and multiple agents scenarios, assuming that the dynamics of each agent is described by a finite-horizon Markov Decision Process of duration T, with *Actions* $\{A_t, t = 1, ..., T-1\}$ and *States* $\{S_t, t = 1..., T\}$, $p(a_1, ..., a_{T-1}, s_1, ..., s_T) = p(s_1)p(a_1)\prod_{t=2}^{T} p(s_t|s_{t-1}, a_{t-1})p(a_t|s_t, a_{t-1})$. Joint distributions on states and actions are characterized with tensors. Forward and backward flows are computed for optimal trajectories.

Probability propagation in Bayesian factor graphs, has been studied in [14], with special attention to factor graphs in Reduced Normal Form (FGrn), that transform a network of random variable nodes to the combination of single-input single-output blocks and replicators.

In [13], the Bayesian model considers a known map with an obstacle avoidance mechanism based on normalization of the conditional transition distribution. The backward distribution plays a crucial role in directing the agents' choices, as it describes a sort of inverse dynamics. The agent always reaches the goal, if a feasible solution exists. The backward distribution is also used for determining the minimum time T to reach the goal. Figure 1 shows the results of a simulation in which various agents move in a train station-like grid. There are counters, rooms, doors, platforms, etc., and each agent has a starting position and a goal. The agents are

*Work partially supported by POR CAMPANIA FESR 2014/2020, ITS for Logistics, awarded to CNIT (Consorzio Nazionale Interuniversitario per le Telecomunicazioni). Research of Pattipati was supported in part by the U.S. Office of Naval Research and US Naval Research Laboratory under Grants #N00014-18-1-1238, #N00173-16-1-G905 and #HPCM034125HQU, and by a Space Technology Research Institutes grant (number 80NSSC19K1076) from NASAs Space Technology Research Grants Program.

†F. A. N. Palmieri with Dipartimento di Ingegneria, Università degli Studi della Campania "Luigi Vanvitelli", Aversa (CE), Italy and visiting at The University of Connecticut, Storrs, CT (USA), francesco.palmieri@unicampania.it

‡K. R. Pattipati is with Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA, krishna.pattipati@uconn.edu

§G. Fioretti, F. Verolla and G. Di Gennaro are with Dipartimento di Ingegneria, Università degli Studi della Campania "Luigi Vanvitelli", Aversa (CE), Italy, giovanni.fioretti, francesco.verolla1@studenti.unicampania.it, giovanni.digennaro@unicampania.it

¶A. Buonanno is with ENEA, Energy Technologies Department, Portici (NA), Italy, amedeo.buonanno@enea.it

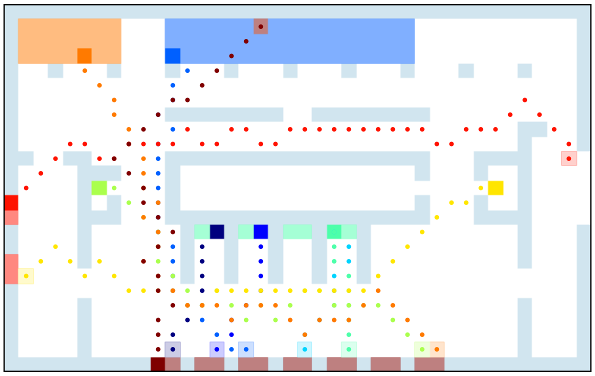

Figure 1. **Agents move in a train station-like environment from their starting positions to their goals**

scheduled, in turn, to make their decisions according to probability messages propagated on the grid. Each agent is seen by the others as an obstacle and the probability flow is recomputed at every time step. The behaviors appear to be quite natural.

2. The limited knowledge scenario

In the current work, we have elaborated on the paradigm and have assumed that the map is not known to the agent a priori, but it is progressively discovered. An agent knows its position and only the part of the map that is able to see, as shown in Figure 2. The exploration/exploitation process is shown in the example of Figure 3. In this scenario, the agent has a 360^o viewing angle and is able to recognize obstacles. At the beginning, the goal is hidden and the agent has to explore the unknown environment via movements, in an attempt to discover it. One of the great features of the probabilistic model is its capability of handling multiple goals simply by spreading the probability distribution (the backward distribution at the end of the chain) across multiple locations [13]. Therefore, in this limited-view scenario, the entire hidden area is set as the goal for the agent, that becomes attracted to it. The backward flow is computed dynamically, and a maximum likelihood algorithm determines the agent's current action. Figure 3 shows how the agent progressively discovers the hidden parts of the map and finally discovers its goal. After that, the steps are the usual ones that lead to the goal.

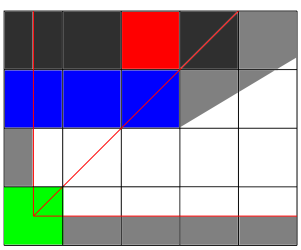

Figure 2. **Agent's viewing geometry. The agent (green) looks at 315° (counterclockwise angles) with 90° viewing aperture. Because of the presence of the obstacles (blue), the goal (red) is not visible. The black areas are the invisible regions. We assume that a partially visible cell is visible (gray).**

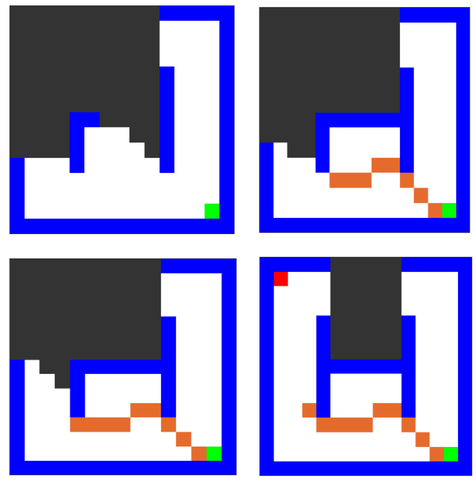

Figure 3. **An agent with limited view that progressively discovers obstacles and goal**

Figure 4 shows a comparison of the paths resulting when an agent has full knowledge of the map, and when it has to explore before seeing the goal. Evidently the path in the first case is much shorter, but the agent achieves the goal in both scenarios. Animation of various simulations are available on our website: https:// www.mlsptlab-unicampania.it/research/topics/path-modeling.html.

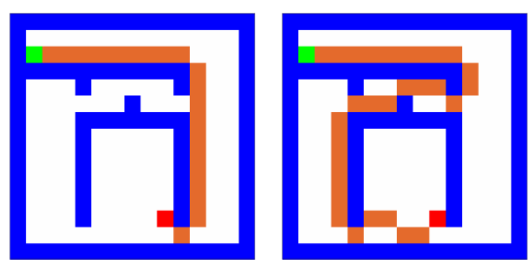

Figure 4. **Paths obtained with a fully known map (left) and after exploration due to limited view (right)**

3. Concluding remarks

We find that the probability propagation paradigm for determining paths in complex scenarios shows great promise and is very robust with respect to map complexity and uncertainties. We are working towards scenarios with multiple agents, the inclusion of reward strategies, application to real maps and to the inclusion of learning.

References

[1] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *arXiv:1905.06113*, 2019.

[2] Pasquale Coscia, Lamberto Ballan, Francesco Palmieri, Alexandre Alahi, and Silvio Savarese, *Linear Artificial Forces for Human Dynamics in Complex Contexts*, pp. 21–33, 01 2020.

[3] P. Coscia, F. Castaldo, F. A. N. Palmieri, A. Alahi, S. Savarese, and L. Ballan, "Long-term path prediction in urban scenarios using circular distributions," *Image and Vision Computing*, vol. 69, pp. 81–91, 2018.

[4] S. Levine, "Reinforcement learning and control as probabilistic inference: Tutorial and review," *arXiv:1805.00909v3 [cs.LG] 20 May 2018*, 2018.

[5] R. Kaplan and K. J. Friston, "Planning and navigation as active inference," *Biological Cybernetics*, vol. 112, pp. 323–347, 2018.

[6] Hagai Attias, "Planning by probabilistic inference," in *Proc. of the 9th Int. Workshop on Artificial Intelligence and Statistics*, 2003.

[7] D. P. Bertsekas, *Reinforcement Learning and Optimal Control*, A. Scientific, Ed., 2019.

[8] Marc Toussaint, "Probabilistic inference as a model of planned behavior," *Künstliche Intelligenz*, vol. 3, 01 2009.

[9] C. L. Buckley, C. S. Kim, S. McGregor, and A. K. Seth, "The free energy principle for action and perception: a mathematica review," *Journal of Mathematica Phychology*, vol. 81, pp. 55–79, 2017.

[10] T. Parr and K. J. Friston, "Generalised free energy and active inference: can the future cause the past?" *BioARXIV*, 2018.

[11] M. Baltieri and C. L. Buckley, "An active inference implementation of phototaxis," *ARXIV*, 2017.

[12] A. Zenon, S. Oleg, and G. Pezzulo, "An information-theoretic perspective on the costs of cognition," *Neuropsychologia*.

[13] Francesco A.N. Palmieri, Krishna R. Pattipati, Giovanni Fioretti, Giovanni Di Gennaro, and Amedeo Buonanno, "Path Planning Using Probability Tensor Flow," *arXiv:2003.02774*, sub. for journal publication, II review, 2020.

[14] Francesco A.N. Palmieri, "A comparison of algorithms for learning hidden variables in bayesian factor graphs in reduced normal form," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2242–2255, 09 2015.