Modular Solvers for Partially Continuous Abstract Markov Decision Processes

Noah Carver, Dr. Cynthia Matuszek, John Winder

Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County

Goal & Motivation

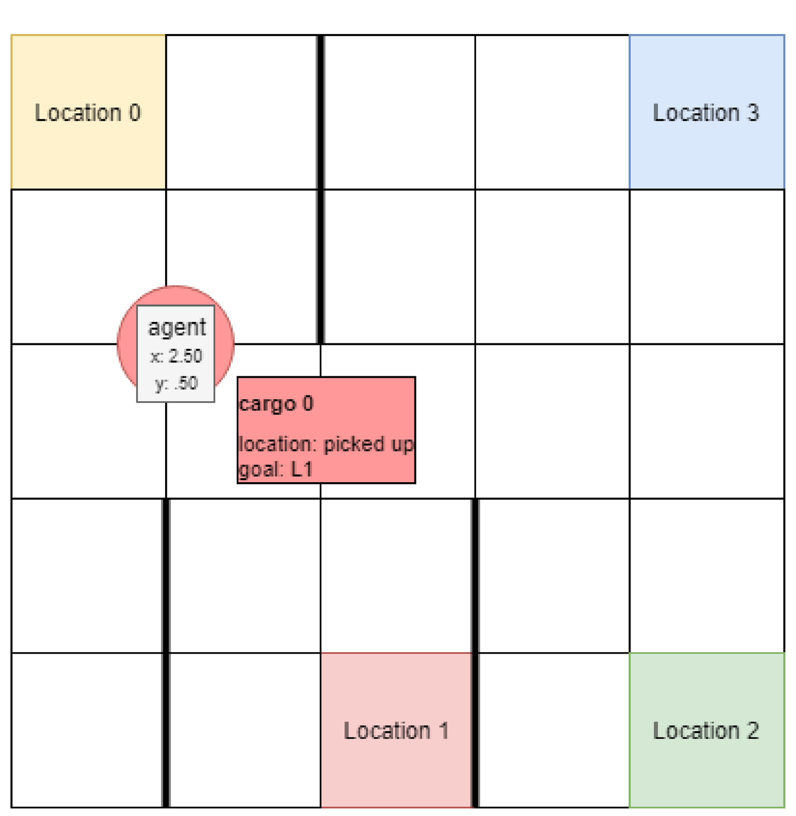
A Formal Structure for the creation and solving of Partially Continuous Abstract Markov Decision Processes.

- Higher level reasoning for continuous problems needs not be continuous.
- More efficient than forcing all subtasks to be solved continuously

We introduce: **Modular Solvers for AMDPs**, a formal way of allowing discrete higher level planning on a continuous metric space.

Background

- **Markov Decision Processes** (MDPs) Phrase problems in ways that reinforcement learners can understand: $\langle \mathcal{S}, \mathcal{A}, \mathrm{T}, \mathrm{R}, \mathcal{E} \rangle$
- **Abstract Markov Decision Processes** (AMDPs) Split problems into a graph of simpler sub-problems as seen in (b): $\langle \bar{\mathcal{S}}, \bar{\mathcal{A}}, \bar{\mathrm{T}}, \bar{\mathrm{R}}, \bar{\mathcal{E}}, \phi \rangle$
- **Continuous vs Discrete**
  - **Discrete:** Like a chessboard. AMDP's are efficient.
  - **Continuous:** Like if Chess had no grid and pieces could go anywhere. AMDP's are not efficient.



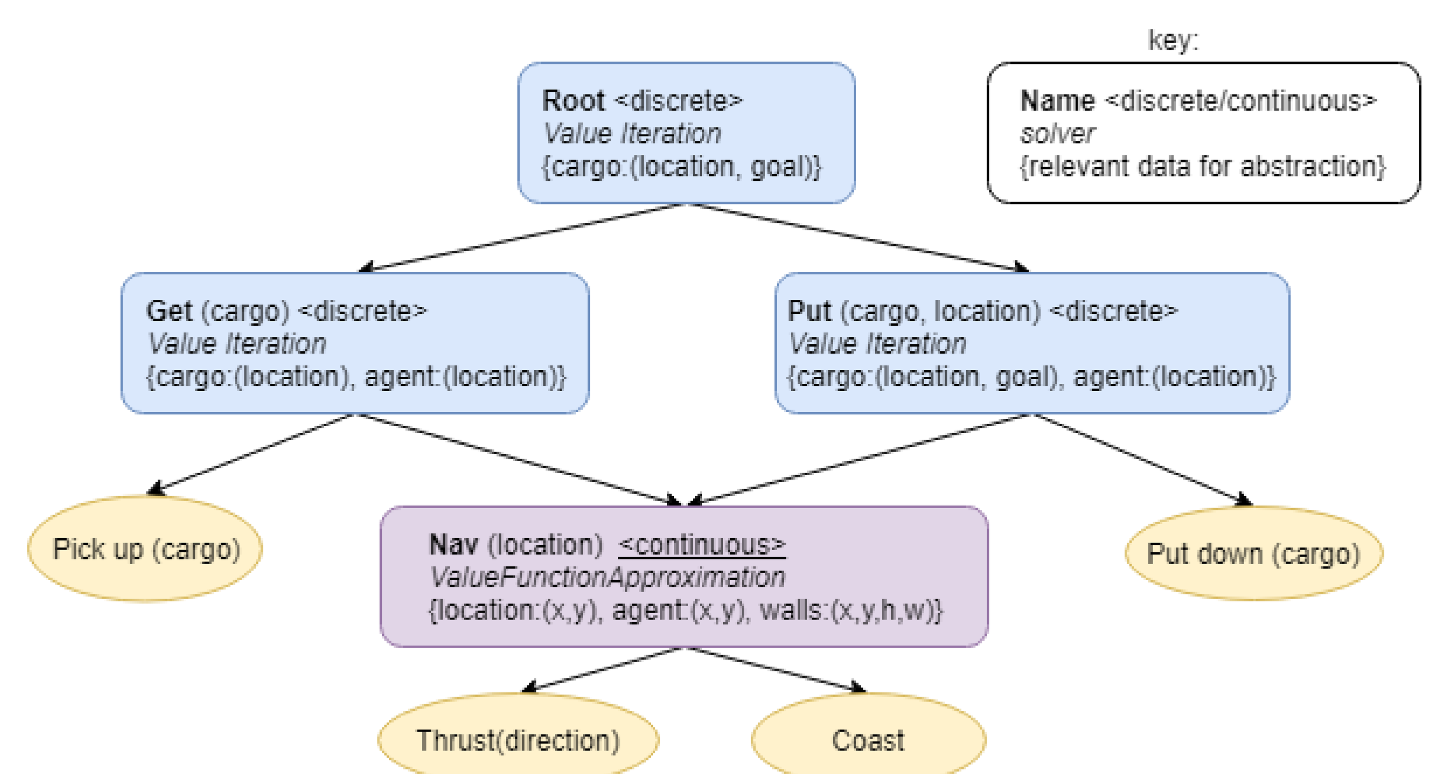
(a) Continuous Domain Visualization

Contact Information

Email: {cmat, jwinder1, ncarver1}@umbc.edu
Web: http://maple.cs.umbc.edu/

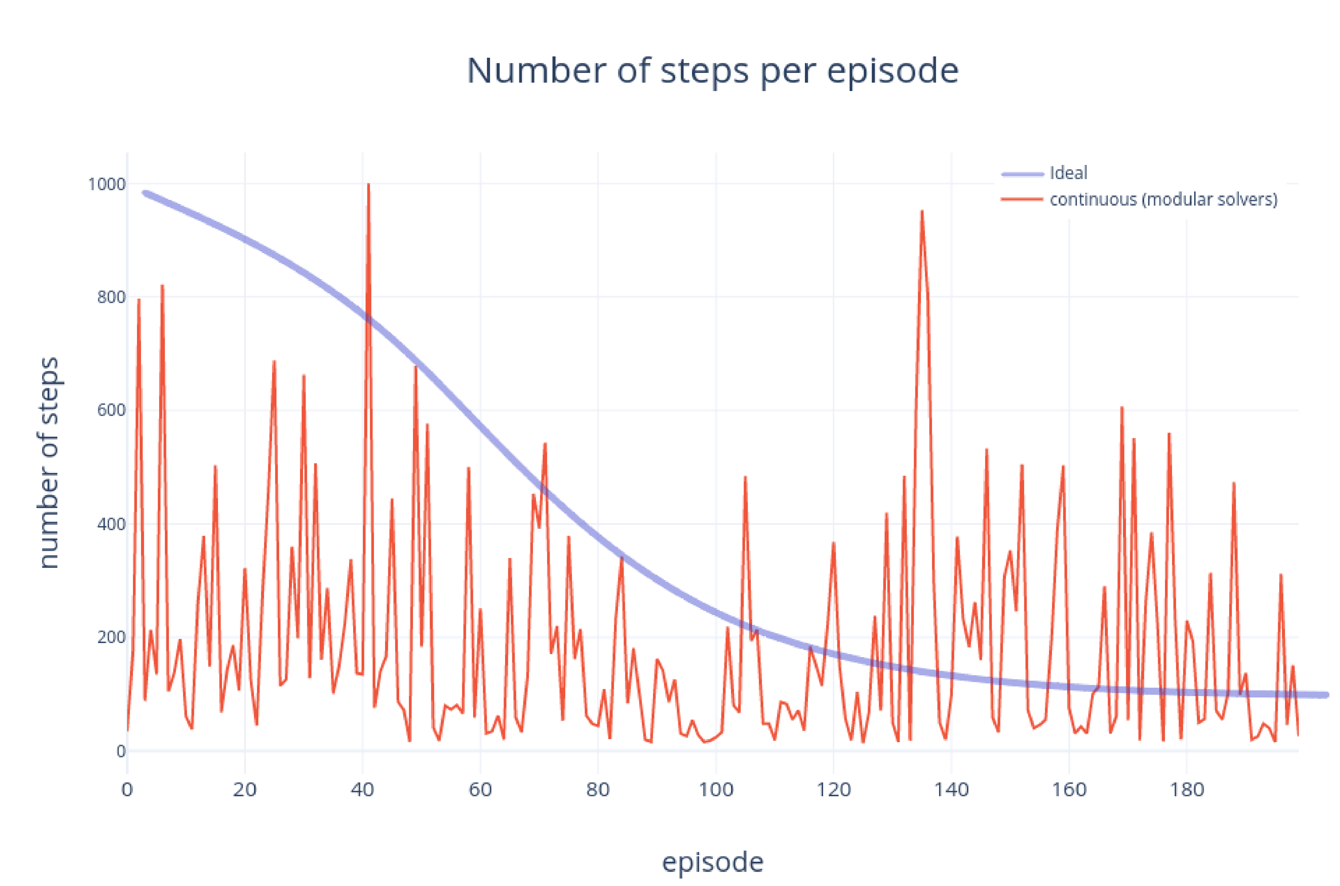
Partially Continuous AMDPs with Modular Solvers

- Each sub-AMDP may be either Continuous or Discrete, so the planner needs to be able to produce policies for both.
- The domain pictured in (a) is a Continuous Taxi domain and was used for all tests.
- Hence, A solver type can be added to $\langle \bar{\mathcal{S}}, \bar{\mathcal{A}}, \bar{\mathrm{T}}, \bar{\mathrm{R}}, \bar{\mathcal{E}}, \phi, \bar{\Pi} \rangle$, where $\bar{\Pi}$ is the solver for each AMDP.

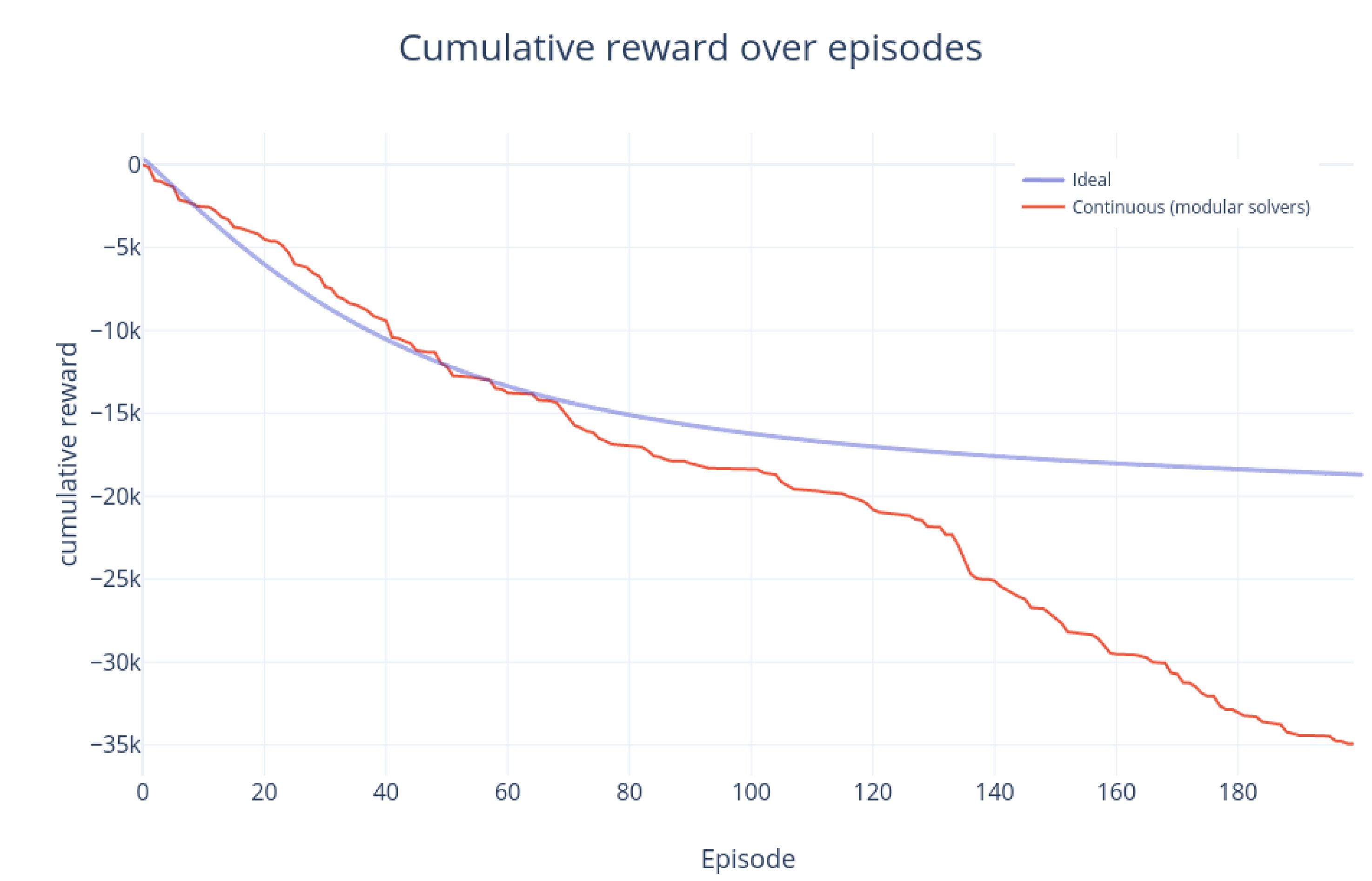


(b) Partially Continuous AMDP with Modular Solvers

Results



(c) Continuous (modular solvers) vs Discrete time steps per episode on standard Taxi



(d) Continuous (modular solvers) vs Discrete reward Per episode on standard Taxi

Conclusion & Future Work

- **Initial performance of solvers on continuous subtasks is lacking**
- Possible cause: **Complexity difference**
  - It takes much longer for the continuous subtask to be solved.
  - This de-sync might be the cause of the jagged episode length.
- Transfer learning may allow Learner to 'focus' on learning continuous subtasks

Acknowledgements

This material is supported by the National Science Foundation under Grant No. IIS-1426452.

References

[1] Thomas G. Dietterich.
Hierarchical reinforcement learning with the MAXQ value function decomposition.
Journal of Artificial Intelligence Research, 13:227–303, 2000.

[2] Neville Mehta.
Hierarchical structure discovery and transfer in sequential decision problems.
PhD thesis, Oregon State University, 2011.

[3] Richard S. Sutton, Doina Precup, and Satinder Singh.
Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning.
Artificial Intelligence, 112(1-2):181–211, 1999.

[4] Nakul Gopalan, Marie desJardins, Michael L. Littman, James MacGlashan, Shawn Squire, Stefanie Tellex, John Winder, and Lawson L.S. Wong.
Planning with abstract Markov decision processes.
In *27th International Conference on Automated Planning and Scheduling*, 2017.