Performance Guarantees for Homomorphisms Beyond Markov Decision Processes

Sultan J. Majeed and Marcus Hutter
Australian National University

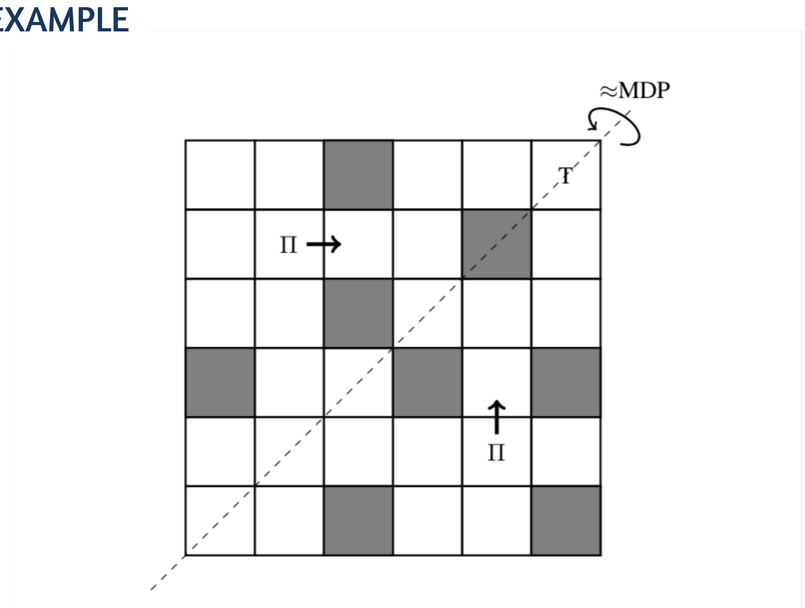
ABSTRACT

Most real-world problems have huge state and/or action spaces. Therefore, a naive application of existing tabular solution methods is not tractable on such problems. Nonetheless, these solution methods are quite useful if an agent has access to a relatively small state-action space homomorphism of the true environment and near-optimal performance is guaranteed by the map. A plethora of research is focused on the case when the homomorphism is a Markovian representation of the underlying process. However, we show that near-optimal performance is sometimes guaranteed even if the homomorphism is non-Markovian.

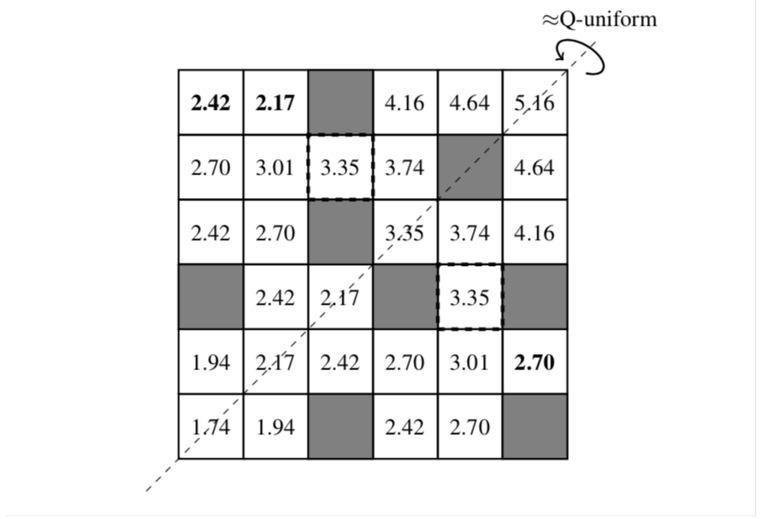
BACKGROUND

- A Reinforcement Learning (RL) agent requires an abstract representation of the underlying environment.
- Typically, this abstraction is assumed to be Markovian.
- The state of the abstraction is sufficient to predict the next state and reward.
- However, it is shown for the state-aggregation case that this Markovian restriction is not necessary for the near-optimal performance.
- A significantly compressed model can be made if we relax the Markovian (i.e. MDP) assumption.
- Can a similar guarantee can be provided for homomorphisms?

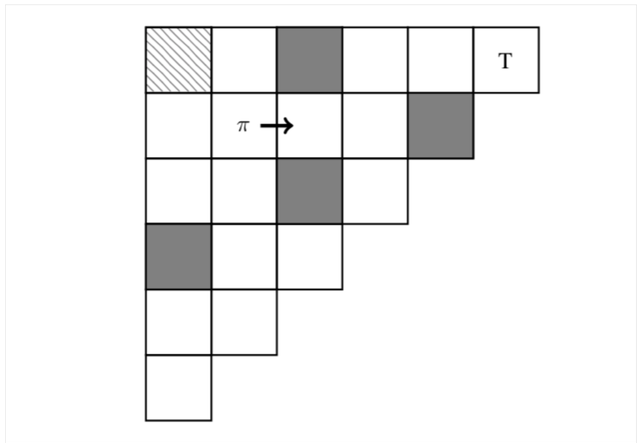
EXAMPLE



A grid-world with an MDP axis of symmetry.



This also has an axis of Q-value uniformity.



The resultant compressed model.

- The above compressed model can be rendered invalid by changing the transition function and rewards, but keeping the Q-uniformity intact. (How? Details are in the full paper.)
- If the states are only considered as Q-uniform, and not MDP then the same model can handle, possibly, infinite grid-words with different transition dynamics.
- Hence, provides a compressed Q-uniform representation for the homomorphisms of this grid-world.
- Does such Q-uniform homomorphism has bounded performance guarantee?

RESULT & CONCLUSION

- Yes, such models do have the performance bound guarantee.
- The performance guarantees can be extended to non-MDP homomorphisms.
- We can aggregate the state-action pairs to a non-MDP map.
- The resultant model still has bounded performance guarantees.
- The agent can have a relatively compressed representation of the underlying environment.

NOTES

This is a theoretical paper with more mathematics than plots. We recommend you to read the full paper for a better understanding of our work.