Using Symmetry in Asymmetrical Markov Decision Processes

[Extended Abstract]

Martin Zinkevich
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
maz+@cs.cmu.edu

Tucker Balch
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
trb+@cs.cmu.edu

ABSTRACT

In Zinkevich and Balch(2001), we proved that a symmetric MDP has a symmetric optimal policy. However, in reality, systems may possess slight asymmetries. In this abstract, we present without proof how in slightly asymmetric MDPs there exists a slightly suboptimal symmetric policy. In order to specify the bound on the suboptimality of this symmetric policy, we use notation used in the other paper.

Consider a world with two agents, Alice and Angie. Alice and Angie are twins which have identical wants, desires, needs, and abilities. If Alice or Angie were in the same spot between an apple orchard and an orange grove, then they would both go to the grove because they both like oranges better. In every situation, Alice and Angie would act identically.

However, suppose that one day Alice realizes that she likes apples a little better than before. Then, if she were in the same spot she was before, she will go to the apple orchard. Going to the orange grove becomes a suboptimal policy. But Angie still likes oranges. The smallest difference in the reward function can result in a situation where there is no optimal symmetric policy.

However, suppose that Alice and Angie wanted to act identically. Then Alice will still eat oranges. If she likes oranges almost as much, then she receives almost as much reward. However, now she will eat oranges every day for the rest of her life. If the change in reward is ΔR, then the suboptimality of her discounted reward would be $\frac{\Delta R}{1-\gamma}$, where γ is the decay constant. In general,

$$\Delta R = \max_{((s,a),(s',a')) \in E_{\mathcal{A}}} \left(R(s,a) - R(s',a') \right)$$

Thus suboptimality increases linearly with the asymmetry of the reward function, the maximum difference in reward between two symmetric actions.

Suppose that instead Alice becomes allergic to oranges. In fact, at any point in time, Alice could sneeze and then the owner of the orange grove will escort her from the orange grove and not allow her to return. Suppose in the interests of maintaining homogeneity, Alice decides to continue to eat oranges, because her chance of sneezing is ΔT, a small value. Thus, at any point in time she has a probability of ΔT of not receiving any more reward, so her discounted expected reward is $\frac{r_{orange}}{1-\gamma(1-\Delta T)}$. This has the same effect as multiplying the decay constant by $(1-\Delta T)$. Suppose $P = \{E(s) | s \in \mathcal{S}\}$, the partition of the states into symmetric sets. In general,

$$\Delta T = \max_{s \in \mathcal{S}, a \in \mathcal{A}} \sum_{p \in P} \left(\left(\sum_{s' \in P} T(s,a,s') \right) - \min_{(s'',a'') \in E_{\mathcal{A}}(s,a)} \sum_{s''' \in P} T(s'',a'',s''') \right).$$

Define

$$|R| = \left(\max_{s \in \mathcal{S}, a \in \mathcal{A}} R(s,a) \right) - \left(\min_{s \in \mathcal{S}, a \in \mathcal{A}} R(s,a) \right).$$

There exists a symmetric policy sym such that for all $s \in \mathcal{S}$:

$$V_{sym}(s) \geq V^*(s) - \frac{\Delta T \gamma |R|}{(1-\gamma)(1-\gamma(1-\Delta T))} - \frac{\Delta R}{1-\gamma}$$

Suppose there was only two actions which had drastically disparate results for Alice and Angie. Suppose only Alice can climb apple trees, and only Angie can climb orange trees. Then if the only fruit left were high in the trees, any symmetric policy would result in drastic results for one agent. Thus, it is the magnitude, and not the number, of deviations from symmetry that determine the suboptimality of the best symmetric strategy.

Acknowledgements

This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. Any opinion, findings, conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

Zinkevich, M. & Balch, T. (2001). Applications of the Theory of Homogeneous Agents To Multiagent Learning. *Eighteenth International Conference on Machine Learning.*

Autonomous Agents 2001 Montreal, Quebec Canada