Speeding Safely: Multi-criteria Optimization in Probabilistic Planning

Michael S. Fulkerson, Michael L. Littman, and Greg A. Keim
Department of Computer Science
Duke University
Durham, North Carolina 27708–0129
{msf,mlittman,keim}@cs.duke.edu

In deterministic planning, an optimal plan reaches the goal in the minimum number of steps. In this work, we find plans for probabilistic domains, in which there is a tradeoff between reaching the goal with high probability (safety) and reaching it quickly (cost).

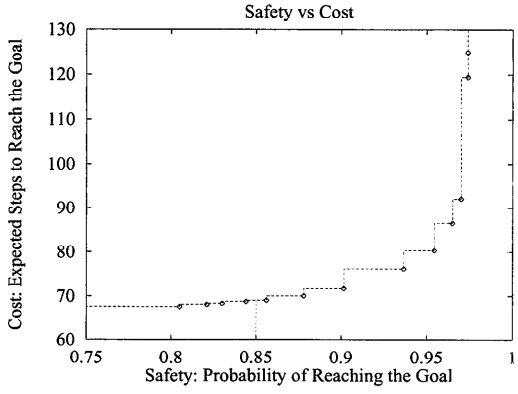
As an example, we consider a large (40463 state) racetrack domain (Barto, Bradtke, & Singh 1995). At any moment, the state of the system is the agent's position in the grid and its two-dimensional velocity. The agent must move through the grid to the finish line without hitting a wall; its actions change its velocity by ± 1 unit in each dimension. With probability 0.2, an action has no change on the velocity (a slip).

The objective is to reach the goal as quickly and safely as possible. We consider three objectives for the agent. The first is to maximize its probability of reaching the goal without crashing (maximize safety), the second is to minimize the expected number of steps to the goal given 100% safety in the face of noise, and the third is to minimize the expected number of steps to the goal given at least 85% safety.

The first two problems can be solved quickly by setting up an appropriate undiscounted Markov decision process (MDP) (Koenig & Simmons 1994). Both MDPs have the same basic form: one state and action for each state and action of the planning problem, an immediate cost for taking a step, a terminal cost for hitting a wall, and a terminal reward for reaching the goal. By varying the costs and rewards, we can find policies with different behaviors. To maximize safety, we set step and wall costs to 0 and goal reward to 1. To minimize steps, we set step cost to 1, wall cost to ∞, and goal reward to 0.

To minimize steps subject to an 85% safety constraint, it is not immediately apparent how the costs and rewards should be defined. We content ourselves with finding a family of policies with different safety properties. For each policy, we measure its safety and its cost (the average number of steps to reach the goal when it is reached). The policy that achieves closest to 85% safety is an approximation of the optimal 85%-safe policy.

A simple algorithm to do this is to repeatedly solve MDPs with step costs varying between 0.0 and 2.8 while holding goal reward and wall costs fixed at 100. The resulting policies, illustrated in the figure, exhibit the safety/cost tradeoff quite nicely; whereas the safest policy takes 126.0 steps on average, the policy closest to 85% safety takes only 69.0 steps. By decreasing safety to 80%, the goal can be reached in 67.5 steps.



The approach described here finesses the problem of guessing the right parameters to achieve desired behavior, but scales badly in domains with more sophisticated cost structures; we plan to address this next.

References

Barto, A. G.; Bradtke, S. J.; and Singh, S. P. 1995. Learning to act using real-time dynamic programming. *Artificial Intelligence* 72(1):81–138.

Koenig, S., and Simmons, R. G. 1994. Risk-sensitive planning with probabilistic decision graphs. *Proc. 4th Int. Conf. on Princ. of KR & Reasoning*, 363–373.