

Robust Trust

Alex Smolin and Piotr Dworczak

Abstract :

An agent chooses an action based on her private information and a recommendation from an informed but potentially misaligned adviser. With a known probability, the adviser truthfully reports his signal; with the remaining probability, he can send any message. We characterize optimal robust decision rules that maximize the agent's worst-case expected payoff. Every optimal rule is equivalent to a trust-region policy in belief space: the adviser's reported beliefs are taken at face value if they fall within the trust region but are otherwise clipped to the trust region's boundary. We derive alignment thresholds above which advice is strictly valuable and fully characterize the solution in both binary-state and binary-action environments.