**深度/宽度搜索解空间，求奖励函数最值**

Base on Markov Decision Process in sequential decision