3

ベルマン方程式

:k回目の更新がなされた価値関数の推定値

4

5

13

【改善】

【評価】

同様に、更新式は

14

3×4のグリッドワールドで考える。

決定論的に遷移するため、

そして、最適方策は最適状態評価関数から求められる。