

多目的マルコフ決定過程のためのアルゴリズム

会員番号01900382 長岡工業高等専門学校 *涌田和芳 WAKUTA Kazuyoshi
長岡工業高等専門学校 外川一仁 TOGAWA Kazuhito

多目的マルコフ決定過程 (MMDP) のためのアルゴリズムは、すでいくつか提案されている。Viswanathan et al.[5] と Novak[3] は、MMDP を多目的LP問題として定式化し、これを解いた。この方法により得られる非支配解の集合は、本講演で述べる最適解の集合より一般に小さい。

White & Kim [6] は、MMDP が特殊な構造をもつ部分的観測MDP と同等であることを述べ、部分的観測MDP のためのアルゴリズムにもとづいた解法を示した。

古川[1] は、割引利得型MMDPのための政策反復法を考案した。Thomas[4] は古川の方法を改良して、平均利得型に応用した。Thomasは完全エルゴード過程の場合を考えたが、伊喜 & 古川[2]は多重連鎖過程の場合の政策反復法について議論した。しかし、これらの政策反復法は、確定的定常政策の中だけでしか議論されていない。

本講演では、割引利得型および平均型MMDPに対して、より広い政策の中で最適な確定的定常政策を求める政策反復法を与える。まず政策反復によって最適である可能性のある候補を絞り、次に最適政策を特徴付ける線形不等式系をLP法で解いて最適性の判定を行う。

参考文献

- [1] N. Furukawa, in: R.Hartley et al. eds., Recent Developments in Markov decision Processes (Academic Press, 1980) pp.205-223.
- [2] T.Iki & N. Furukawa, Memoirs of the Faculty of Education, Miyazaki University, 54.55(1984)1-10.
- [3] J.Novak, Optimization 20(1989)651-670.
- [4] L.C.Thomas, in: S.French et al. eds., Multi-Objective Decision Making(Academic Press, 1983) pp.77-94.
- [5] B.Viswanathan et al., TIMS Studies in the Management Sciences 6(1977)263-272.
- [6] C.C.White & K.W.Kim, Large Scale Systems 1(1980)129-140.