

多目的マルコフ決定過程のためのプログラムソフト

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

1. はじめに

前回のOR学会で、多目的マルコフ決定過程(MMDP)の解法のためのアルゴリズムについて報告した。今回、そのソフトウェアの開発を行ったので報告する。

MMDPのためのアルゴリズムは、すでにいくつか提案されている。

(1) Viswanathan et al.[5], Novak[3]は、MMDPを多目的LP問題として定式化してこれを解いた。

(2) 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。

(3) 古川[1], Thomas[4], 伊喜&古川[2]は、MMDPのための政策反復法を考察し、確定的定常政策の中から最適なものを求めた。

(4) 涌田[7]は、(3)の政策反復法を改良し、一般の政策の中から最適な確定的定常政策を求めた。最初に政策反復によって最適政策の候補を絞り、次に最適性を特徴づける線形不等式系をフーリエ消去法で解いて最適解を求めた。

(5) 涌田&外川[8]は、涌田[7]の線形不等式系をLP法で解くことにより、アルゴリズムを実行可能なものとした。

MMDPのためのアルゴリズムとしては、上記の多目的LP法と(多目的版)政策改良法が有効である。涌田&外川[8]は、従来のアルゴリズムにおける次の欠点を解消している。

・多目的LP問題として定式化する場合は、正の初期分布を与える必要がある。しかし、これによって得られる非支配解の集合は通常のMMDPの解より小さい集合である。我々の方法は、すべての初期状態について最適な通常のMMDPの解を得ることができる。

・これまでの政策改良法では、上述のように確定的定常政策の中だけから最適解を求めている。したがって、一般の政策を考慮に入れると最適解でないものも解として得られる可能性がある。我々の方法では、それは解から排除される。

2. アルゴリズム

Phase I (政策反復)

$E_n, n \geq 0$: n 回の政策反復を通して、他の政策に支配されない政策の集合

$F_n, n \geq 0$: n 回の政策反復を通して、他の政策に支配される政策の集合

とする。

1. $E_0 = F_0 = \phi$ とおき、 $f_1 \in \Pi_D$ を任意に選ぶ。

2. $f_n \in \Pi_D$ に対して、 $I_{f_n}(i), i \in S$ を求める。

$$3. A_{f_n} = \{g \in (\Pi_D \setminus (E_{n-1} \cup F_{n-1})) \mid I_{f_n}^g - I_{f_n}^{f_n} \geq 0\}$$

$$B_{f_n} = \{g \in (\Pi_D \setminus (E_{n-1} \cup F_{n-1})) \mid I_{f_n}^g - I_{f_n}^{f_n} \leq 0\}$$

$$C_{f_n} = \{g \in (\Pi_D \setminus (E_n \cup F_n)) \mid I_{f_n}^g - I_{f_n}^{f_n} = 0\}$$

を定める.

4. $A_{f_n} \neq \phi$ ならば,

$$E_n = E_{n-1}, F_n = F_{n-1} \cup B_{f_n} \cup C_{f_n}$$

とおく. $f_{n+1} \in A_{f_n}$ を選び, Step 2へ戻る.

$A_{f_n} = \phi$ ならば,

$$E_n = E_{n-1} \cup C_{f_n}, F_n = F_{n-1} \cup B_{f_n}$$

とおく. $f_{n+1} \in (\Pi_D \setminus (E_n \cup F_n))$ を選び, Step 2へ戻る.

5. $E_n \cup F_n = \Pi_D$ となったら手続きを停止する. $E^* = E_n$ において Phase IIへ行く.

Phase II (最適性の判定 (a))

E :最適な確定的定常政策の集合とおく.

1. 各 $g \in E^*$ に対して $LP(S_0)$ を解いて最適かどうか判定する. その結果, E の部分集合 E' がもとまる.

Phase III (最適性の判定 (b))

1. 各 $g \in (E^* \setminus E')$ に対して $LP(U_1), \dots, LP(U_N)$ を解いて最適かどうか判定する. その結果 E がもとまる.

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, Mem. Fac. Educ. 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 Sci. 6(1977)263-272.
- [6] C. C. White & K. W. Kim, Large Scale Systems 1(1980)129-140.
- [7] K. Wakuta, Vector-valued Markov decision processes and the systems of linear inequalities, Stochastic Processes and their Applications 56(1995) 159-169.
- [8] K. Wakuta & K. Togawa, Algorithms for vector-valued Markov decision processes, Submitted.