DETERMINAREA STRATEGIILOR OPTIME ÎN PROCESELE MARKOV DECIZIONALE ŞI SOLUŢIONAREA JOCURILOR STOCASTICE DINAMICE POZIŢIONALE

Igor MANDRIC, Facultatea de Matematică şi Informatică

The aim of the study is to research methods for determining optimal strategies in Markov decision processes and elaborating an iterative numerical algorithm for determining Nash equilibria in stochastic dynamic positional games.

Procesele Markov decizionale sunt foarte răspândite în modelarea luării de decizii secvențiale, problema care apare frecvent în diferite domenii ale ştiinţelor inginereşti, cercetărilor operaționale, economiei, ştiinţelor sociale. Însă este cunoscut faptul că, în mai multe aplicaţii reale, care pot folosi conceptul de procese Markov decizionale, există un număr mare de stări şi în fiecare stare sunt posibile mai multe acțiuni. Ca în consecință, simpla metodă de enumerare a tuturor situaţiilor posibile devine lipsită de sens. Pentru soluționarea efectivă a problemelor decizionale, este nevoie de nişte algoritmi puternici, care ar putea să furnizeze soluțiile optime în timp rezonabil.

În lucrarea dată, se studiază procesele Markov decizionale; sunt consideraţi algoritmi existenți pentru determinarea strategiilor optime ale proceselor Markov decizionale; sunt descrise diferite criterii de optimalitate, însă ca criterii de bază s-au studiat criteriile costului total mediu (precum şi acel discontat) şi a costului mediu la un pas de trecere. Este definit un tip nou de jocuri, care reprezintă un proces Markov decizional, dirijat de mai mulți decidenți. Astfel de jocuri se

numesc *jocuri stocastice dinamice poziţionale*. Aceste jocuri se studiază sub aspectul strategiilor optime, adică sub aspectul determinării situațiilor de echilibru în sensul Nash. A fost argumentată existența echilibrelor Nash, precum şi au fost elaborați algoritmi numerici noi pentru determinarea echilibrelor Nash. Baza teoretică a algoritmilor se sprijină pe conceptul iterațiilor după strategii (*policy iteration*), introdus de Howard în anul 1960 [1].

Definim jocul [2, 3]: fie că avem un proces Markov decizional *(X, A, p, c)*, unde:

1) mulțimea de stări X este finită;

2) mulțimea acțiunilor *A* în fiecare stare este o mulțime finită;

3) *p* este funcția probabilităților de tranziție;

4) *c* este funcția costului de tranziție.

Mulțimea de stări ale sistemului este divizată între jucători, adică mulțimea *X* este reprezentată ca o reuniune a *m* mulțimi distincte:

$$X = X_1 \cup X_2 \cup ... \cup X_m$$

Fiecare jucător în pozițiile ce îi aparţin fixează câte o acțiune. Noi considerăm că strategiile sunt staționare, din această cauză fiecare jucător fixează în fiecare stare a lui o acțiune ce nu va depinde de timp (de epoca decizională) .

$$s^1 : x \to a, a \in A^1 \text{(}x\text{)}, pentru x \in X^1,$$
$$s^2 : x \to a, a \in A^2 \text{(}x\text{)}, pentru x \in X^2,$$
$$...$$
$$s^m : x \to a, a \in A^m \text{(}x\text{)}, pentru x \in X^m,$$

unde $A^i(x)$ reprezintă mulțimea de acțiuni ale jucătorului i în starea $x \in X^i$.

Jucătorii, independent, îşi fixează acțiunile, adică îşi aleg strategiile lor. În urma acestei alegeri, se creează o situație de joc:

$$s = (s^1, s^2, ..., s^m)$$

Dacă este dată starea inițială, noi putem determina pentru fiecare jucător costul lui mediu la un pas de trecere, având în vedere matricea costurilor c^i_{xy} . Deci, funcția de câştig a fiecărui jucător se defineşte astfel:

$$F^i_{x_0}(s) = F^i_{x_0}(s^1, s^2, ..., s^m) = \sigma^i_{x_0}(s^1, s^2, ..., s^m),\ i = 1,2,..,m$$

Algoritmul iterativ pentru jocul cu *n* jucători (proces Markov perfect) cu criteriul costului mediu la un pas de trecere

Pasul *0*. Fixăm strategiile arbitrare:

$$s_0^i : x \to y, y \in X(x), pentru\, x \in X^i, i = 1, 2, ..., n$$

Pasul *k*. Calculăm:

$$\mu_{x,s_{k-1}} = \sum_{y \in X(x)} p_{x,y}^{s_{k-1}} c_{x,y}^{s_{k-1}}, \forall x \in X_i, i = 1, 2, ..., n$$

Rezolvăm sistemul de ecuații:

$$\sigma_x = \mu_{x,s_{k-1}(x_i)} + \gamma \sum_{y \in X(x)} p_{x,y}^{s_{k-1}} \sigma_y, x \in X_i, i = 1, 2, ..., n$$

Determinăm:

$$s_k^i(x) \in \underset{a \in A(x)}{arg} \left\{ \mu_{x,a} + \gamma \sum_{y \in X(x)} p_{x,y}^a \sigma_y^{k-1} \right\}, \forall x \in X_i, i = 1, 2, ..., n$$

Verificăm, dacă $s_k^i = s_{k-1}^i, \forall x \in X_i$. Dacă e așa, atunci STOP și setăm $s^{1*} = s_{k-1}^1, s^{2*} = s_{k-1}^2, ..., s^{n*} = s_{k-1}^n$, altfel trecem la pasul *k+1*.

Convergența acestui algoritm poate fi argumentată într-un fel asemănător cu convergența algoritmului iterativ după strategii în procesele Markov decizionale [4, 5].

Referințe:

1. HOWARD, R.A. *Dynamic Programming and Markov Processes.* Wiley, 1960.
2. LOZOVANU, D. The game-theoretical approach to Markov decision problems and determining Nash equilibria for stochastic positional games. In: *Int. J. Mathematical Modelling and Numerical Optimization.* 2011, no.2 (2), p.162-164.
3. LOZOVANU, D., PICKL, S. *Determining optimal stationary strategies for discounted stochastic optimal control problem on networks.* 9th Cologne-Twente Workshop on Graphs and Combinatorial Optimization, Cologne (U.Figle,R.Shrader, D. Herrmann eds.), 2010, p.115-118.
4. PUTERMAN, M. *Markov Decision Processes: Stochastic Dynamic Programming*. New York: John Wiley, 1994.
5. SHAPLEY, L.S. *Stochastic Games*. Princeton University, 1953.

Recomandat
Dmitrii LOZOVANU, dr. hab., prof. univ.,