Доклады Академии наук СССР
1974. Том 215, № 6

УДК 62-50:518.9 МАТЕМАТИКА

Е. Т. ГУРВИЧ

О ЦЕЛЯХ АДАПТИВНОГО УПРАВЛЕНИЯ, ДОСТИГАЕМЫХ КОЛЛЕКТИВОМ АВТОМАТОВ

(Представлено академиком А. А. Дородницыным 18 IX 1973)

Рассматриваются автоматные системы управления ([1]) для одного класса управляемых случайных процессов.

1. Пусть заданы измеримое фазовое пространство $(X, \mathscr{B})$ и пространство управлений Y. Управляемый процесс ξ_t будем называть однородным процессом с независимыми значениями (о.п.н.з.), если он характеризуется системой условных вероятностей

$$\mu(B|y)=P\{\xi_{t+1}\in B|y(t)=y\},\quad B\in\mathscr{B}.$$

Мы будем рассматривать случай, когда $X=X^{(1)}\times X^{(2)}\times\ldots\times X^{(N)}$, $Y=Y^{(1)}\times Y^{(2)}\times\ldots\times Y^{(N)}$ и для каждого множества $B\in\mathscr{B}$ $B=B^{(1)}\times B^{(2)}\times\ldots\times B^{(N)}$, $B^{(i)}\subset X^{(i)}$ и $\mu(B|y)=\prod\limits_{i=1}^{N}\mu_i(B^{(i)}|y)$. Значения процесса и управляющей переменной являются при этом векторными величинами одинаковой размерности N: $\xi_t=(\xi_t^{(1)},\ldots,\xi_t^{(N)})$, $\bar{y}(t)=(y^{(1)}(t),\ldots,y^{(N)}(t))$. В дальнейшем будем считать все множества $X^{(i)}$ состоящими из двух элементов* $\{0, 1\}$, а множества $Y^{(i)}$ — конечными.

2. Примем систему управления в виде прямого произведения N одинаковых автоматов $L=A_1\times\ldots\times A_N$. В моменты времени $t=1, 2, \ldots$ автомат A_i выбирает значение i-й компоненты управляющего вектора $\bar{y}(t)$, а на его вход поступает величина ξ_t. При таком способе организации системы L она является децентрализованной.

Автоматы $A_1,\ldots,A_N$ мы будем брать из класса C, к которому отнесем автоматы $D_{\varkappa,n}$ ([2]), $K_{\varkappa,n}$ ([3]), $V_{\varkappa,n}$ ([4]) ($\varkappa$ означает число выходных сигналов, а n — глубину памяти автомата). Если задана последовательность $\{L_n\}$, $n=1, 2, \ldots$, управляющих систем такая, что $L_n=A_1{}^n\times\ldots\times A_N{}^n$, причем автоматы $A_i{}^n$ принадлежат классу C и образуют последовательность, асимптотически оптимальную в любой стационарной среде ([5]), то мы будем называть $\{L_n\}$ последовательностью типа C.

3. Управляющие системы L_n являются «обучаемыми системами» и для некоторых целей управления — «адаптивными системами» ([1]). Ставится задача выяснить, какие цели достигаются ими при управлении процессами охарактеризованного класса. Совокупность $L_n\otimes\xi_t$ обучаемой системы и управляемого процесса описывается марковской цепью M_n, состояниями которой являются наборы $\bar{s}\in S^n$ состояний всех автоматов. Множество S^n разбивается на непересекающиеся подмножества $S_k{}^n$ такие, что из $\bar{s}(t)\in S_k{}^n$ следует $\bar{y}(t)=\bar{y}_k$.

---

* Приводимые далее результаты можно распространить и на случай произвольных ограниченных множеств, если значения $x_i\in[A_i, B_i]$ каждой компоненты процесса подавать сначала на генератор случайных чисел, выдающий число 0 или 1 с вероятностями $(B_i-x_i)/(B_i-A_i)$ и $(x_i-A_i)/(B_i-A_i)$ соответственно.

Для каждого значения управляющего вектора введем его финальную вероятность $\pi(\bar{y}_k, n) = \sum_{\bar{s} \in S_k^n} \pi^{M_n}(\bar{s})$, где $\pi^{M_n}(\bar{s})$ — финальная вероятность состояния $\bar{s}$ в цепи M_n.

Определим предельный средний выигрыш автомата A_i как

$$w_n^{(i)} = \sum_{\bar{y} \in Y} \pi(\bar{y}, n) f_i(\bar{y}),$$

где $f_i(\bar{y}) = E(\xi^{(i)} | \bar{y})$ — математическое ожидание i-й компоненты процесса при значении $\bar{y}$ управляющей величины, и будем считать, что целью управления является максимизация некоторой функции (в дальнейшем мы всюду будем считать ее линейной) от средних выигрышей всех автоматов $F(\bar{w}_n) = F(w_n^{(1)}, \ldots, w_n^{(N)})$.

В качестве такой функции может быть взята, например, сумма компонент $F(\bar{w}_n) = \sum_{i=1}^{N} w_n^{(i)}$, минимальная из компонент $F(\bar{w}_n) = \min_{1 \leqslant i \leqslant N} w_n^{(i)}$ и т. д. Обозначим

$$\bar{F} = \max_{\bar{y} \in Y} F[f_1(\bar{y}), \ldots, f_N(\bar{y})]$$

и назовем систему L оптимальной, если $F(\bar{w}) = \bar{F}$, и ε-оптимальной, если $F(\bar{w}) \geqslant \bar{F} - \varepsilon$.

4. Выделим в каждом из подмножеств S_k^n состояние φ_k, которому соответствуют состояния максимальной глубины всех автоматов. Процесс M_n, рассматриваемый лишь в моменты попадания в множество $\Phi = \{\varphi_1, \varphi_2, \ldots\}$, образует новую цепь Маркова G_n, финальное распределение которой обозначим $\{\pi^{G_n}\}$.

Лемма 1. *Пусть задана последовательность $\{L_n\}$ управляющих систем типа С. Тогда для любого о.п.н.з. ξ_t существуют константы $c_1 > 0$, $c_2 > 0$ такие, что при всех n и k*

$$c_1 \pi^{G_n}(\varphi_k) \leqslant \pi(\bar{y}_k, n) \leqslant c_2 \pi^{G_n}(\varphi_k).$$

Лемма 1 сводит исходную задачу к изучению последовательности марковских цепей $\{G_n\}$, в которой, в отличие от последовательности $\{M_n\}$, число состояний не растет с увеличением n.

Обозначим $T(a, A_i^n)$ среднее время до смены действия автоматом A_i^n в стационарной среде с математическим ожиданием выигрыша a. Для переходных вероятностей $p_{kl}^{G_n}$ цепи G_n удается получить следующую оценку.

Лемма 2. *Если величины $\bar{y}_k \in Y$ и $\bar{y}_l \in Y$ отличаются компонентами $i_1, \ldots, i_q$, то существуют константы $c_3 > 0$, $c_4 > 0$ такие, что при всех n*

$$c_3 \prod_{r=1}^{q} (T[f_{i_r}(\bar{y}_k), A_{i_r}^n])^{-1} \leqslant p_{kl}^{G_n} \leqslant c_4 \sum_{i=1}^{N} (T[f_i(\bar{y}_k), A_i^n])^{-1}.$$

5. Будем ставить в соответствие каждому процессу ξ_t ориентированный граф V следующим образом: 1) множество вершин V изоморфно множеству Y, 2) из k в l существует дуга, если $\bar{y}_k$ и $\bar{y}_l$ отличаются только одной (ν-й) компонентой и $f_\nu(\bar{y}_k) = \min_{1 \leqslant i \leqslant N} f_i(\bar{y}_k)$.

Теорема 1. *Пусть задана последовательность $\{L_n\}$ обучаемых систем типа С. Если о.п.н.з. ξ_t таков, что множество $U \subset Y$, на котором до-*

стигается $\max\limits_{\bar y \in Y} \min\limits_{1\leqslant i\leqslant N} f_i(\bar y)$, *достижимо из любой вершины* V, *то для целевой функции «максимум минимальной компоненты»* $\lim\limits_{n\to\infty} F(\bar w_n)=\bar F$.

Теорема означает, что для любого $\varepsilon>0$ существует $n(\varepsilon)$ такое, что, если $n>n(\varepsilon)$, система L_n является ε-оптимальной.

Пример 1. $N=2$, $Y^{(1)}=\{1, 2\}$, $Y^{(2)}=\{1, 2, 3\}$; $f_1(1, 1)=f_1(1, 3)=$ $=f_2(2, 1)=f_1(2, 2)=0{,}9$; $f_2(2, 3)=0{,}5$; $f_1(2, 3)=0{,}4$; $f_2(1, 1)=$ $=f_1(2, 1)=f_2(2, 2)=0{,}3$; $f_2(1, 2)=f_2(1, 3)=0{,}2$; $f_1(1, 2)=0{,}1$; граф V для рассматриваемого процесса показан на рис. 1. Множество U состоит из вершины (2, 3), которая достижима из любой вершины графа; $\bar F=\max\limits_{\bar y}\min\limits_{i} f_i(\bar y)=0{,}4$. В силу теоремы 1 получаем, что $\lim\limits_{n\to\infty} F(\bar w_n)=0{,}4$ для любой последовательности управляющих систем типа C.

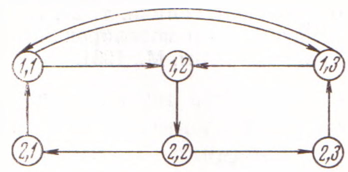

Рис. 1

6. Рассмотрим теперь процессы, у которых при любом управлении математическое ожидание всех компонент одинаково:

$$f_1(\bar y)=f_2(\bar y)=\ldots=f_N(\bar y)\quad \forall \bar y\in Y. \qquad (1)$$

Теорема 2. *Пусть задана последовательность управляющих систем типа* C *и о.п.н.з.* ξ_t, *для которого выполняется* (1).

Тогда для любой линейной целевой функции F, *не убывающей по всем аргументам*, $\lim\limits_{n\to\infty} F(\bar w_n)=\bar F$.

Пример 2. N, $Y^{(1)}$, $Y^{(2)}$ — как в примере 1, $f_1(1, 1)=f_2(1, 1)=0{,}5$; $f_1(1, 2)=f_2(1, 2)=0{,}6$; $f_1(1, 3)=f_2(1, 3)=0{,}3$; $f_1(2, 1)=f_2(2, 1)=0{,}7$; $f_1(2, 2)=f_2(2, 2)=0{,}4$; $f_1(2, 3)=f_2(2, 3)=0{,}6$.

Если задана целевая функция $F(\bar w)=w^{(1)}+w^{(2)}$, то $\bar F=1{,}4$ и в силу теоремы 2 $\lim\limits_{n\to\infty} F(\bar w_n)=1{,}4$.

7. Взаимодействие L и ξ_t можно также интерпретировать как «игру автоматов» [5], при этом каждому значению управляющего вектора $\bar y$ соответствует некоторая партия игры, величина $f_i(\bar y)$ имеет смысл среднего выигрыша i-го участника в партии $\bar y$, $\pi(\bar y, n)$ — финальная вероятность партии $\bar y$ и т. д.

Все результаты, полученные для управления векторными о.п.н.з., могут быть переформулированы в терминах игр автоматов. Мы, однако, не будем делать этого, а дадим только игровую интерпретацию приведенных выше примеров.

Оба примера соответствуют играм двух автоматов, первый из которых имеет два, а второй — три действия. Из теоремы 1 следует, что если в игре участвуют автоматы из последовательностей $\{D_{2, n}\}$ и $\{D_{3, n}\}$, $\{K_{2, n}\}$ и $\{K_{3, n}\}$, то финальная вероятность партии (2, 3) стремится к единице: $\pi((2, 3), n)\to 1$, $n\to\infty$. При рассмотрении игр автоматов, как правило, выделяют «ситуации равновесия» (или «партии Нэша») [6], в которых ни одному из участников невыгодно менять стратегию, если стратегии остальных игроков не меняются. Множество ситуаций равновесия, вообще говоря, не совпадает с фигурирующим в теореме 1 множеством U «максминных» партий; так, в примере 1 существует единственная ситуация равновесия (1, 1), однако автоматы разыгрывают не ее, а максминную партию (2, 3). Лишь при некоторых специальных условиях эти два множества совпадают и тогда теорема 1 дает достаточное условие разыгрывания автоматами ситуации равновесия.

В игре, соответствующей примеру 2, из теоремы 2 вытекает, что $\lim\limits_{n\to\infty} \pi((2, 1), n)=1$, т. е. при увеличении глубины памяти начинает разыгрываться исключительно партия (2, 1), в которой выигрыш обоих автоматов максимален.

Автор выражает глубокую признательность В. Г. Сраговичу за постановку задачи и большую помощь в работе.

Московский физико-технический институт
Долгопрудный Московской обл.

Поступило
6 IX 1973

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] *В. Г. Срагович*, Техническая кибернетика, № 4, 114 (1972). [2] *H. Robbins*, Proc. Nat. Acad. Sci. U. S. A., v. 42, № 12, 920 (1956). [3] *В. Ю. Крылов*, Автоматика и телемеханика, т. 24, № 9, 1227 (1963). [4] *Н. П. Канделаки, Г. Н. Церцвадзе*, Автоматика и телемеханика, т. 27, № 6, 115 (1966). [5] *М. Л. Цетлин*, Исследования по теории автоматов и моделированию биологических систем, М., 1969. [6] *Дж. Нэш*, В сборн. Матричные игры, М., 1961.