

**А. В. Колногоров** (Великий Новгород, НовГУ). **Стратегии в задаче о многоруком бандите для систем с параллельной обработкой данных.**

Многорукий бандит математически описывается как управляемый случайный процесс  $\xi_t$ ,  $t = 1, \dots, T$ , принимающий значения 1, 0 с вероятностями  $p_\ell$ ,  $q_\ell$  соответственно, интерпретируемые как доходы и зависящие только от выбираемых в текущие моменты времени вариантов  $\ell = 1, \dots, K$ . Вероятности  $p_1, \dots, p_K$  предполагаются неизвестными, а цель управления формулируется в минимаксной постановке и, неформально, означает максимизацию (в некотором смысле) математического ожидания полного дохода. При этом обычно предполагается, что данные обрабатываются последовательно, по одному. Однако существует много ситуаций, когда такой подход неприемлем. Ниже предлагаются две стратегии, допускающие параллельную обработку данных.

*R-этапная стратегия.* Она состоит в том, что управление разбивается на небольшое число этапов, на которых применение вариантов осуществляется параллельно. Пусть, например, требуется лечить большую группу, состоящую из  $T$  пациентов, и для лечения имеются  $K$  альтернативных лекарств различной эффективности. Тогда на первом этапе каждое из лекарств может быть дано группе из  $\tau_1$  пациентов. Далее результаты сравниваются, и наиболее эффективное лекарство на заключительном этапе дается оставшимся  $T - K\tau_1$  пациентам. При этом продолжительность лечения определяется не числом пациентов, а длительностью двух этапов лечения. В случае  $r$ -этапной стратегии сравнение вариантов и последовательное отбрасывание худших из них происходит на начальных этапах общим числом не более  $(r - 1)$ . На  $i$ -м этапе ( $i = 1, \dots, r - 1$ ) тестируемые варианты применяются к тестовым партиям объема  $\tau_i$ , затем сравниваются полные доходы за каждый из вариантов и те варианты, которым соответствуют доходы, отличающиеся от максимального на данном этапе по крайней мере на величину порога  $\Delta_i$ , отбрасываются из группы. Если в результате оказываются отброшенными все варианты кроме одного, то происходит переход к заключительному этапу, на котором применяется только этот вариант. Предполагается, что  $\Delta_{r-1} = 0$ , поэтому переход к заключительному этапу после  $(r - 1)$ -ого этапа происходит в обязательном порядке. Асимптотически оптимальные (при  $T \rightarrow \infty$ ) оценки для тестовых групп и порогов  $\{\tau_i, \Delta_i\}$ ,  $i = 1, \dots, r - 1$  представлены в [1].

*Пороговая стратегия.* В этом случае также выполняется отбрасывание худших вариантов, но оно происходит сразу после достижения порога. Рассмотрим сначала процедуру отбрасывания одного «худшего» варианта. Она состоит в том, что все тестируемые варианты применяются последовательно до тех пор, пока разность полных доходов за какие-либо два из них не превысит заданного порога или не истечет время управления; если время управления не истекло, то вариант, которому соответствует наименьший полный доход, отбрасывается из группы. Эту процедуру следует последовательно применять, начиная с группы из  $K$  вариантов, пока не истечет время управления или не останется один вариант, который и будет применяться до конца. Значения порогов зависят от количества вариантов в группе, а накопленные доходы могут сохраняться после отбрасывания очередного варианта, или обнуляться. При этом пороговая стратегия обладает следующим важным свойством [2]. Рассмотрим целое число  $m$ , характеризующее количество данных, которые можно обрабатывать параллельно и предположим для простоты, что  $T$  кратно  $m$ . Смену вариантов разрешим только в моменты времени кратные  $m$ , т. е.  $m, 2m, 3m, \dots$ . Вместо процесса  $\xi_t$ ,  $t = 1, \dots, T$ , будем управлять процессом  $\eta_n = \sum_{t=(n-1)m+1}^{nm} \xi_t$ ,  $n = 1, \dots, T/m$ , используя пороговую стратегию. Если можно использовать параллельную обработку, то время функционирования системы уменьшается в  $m$  раз, при этом величина минимаксного риска асимптотически (при  $T \rightarrow \infty$ ) остается неизменной.

## СПИСОК ЛИТЕРАТУРЫ

1. *Kolnogorov A. V., Melnikova S. V.* Minimax  $r$ -stage strategy for the multi-armed bandit problem. Prepr. 9th IFAC Workshop "Adaptive Systems in Control and System Processing" (ALCOSP'07), Saint-Petersburg, 2007 (6 pages).
2. *Колмогоров А. В., Шелонина Т. Н.* Пороговая стратегия управления в случайной среде. — Труды VI Международной конференции «Идентификация систем и задачи управления» SICPRO'07. М.: ИПУ РАН, 2007, ISBN-5-201-14992-8, с. 1500–1512.