

УДК 519.865

ОБОБЩЕНИЕ ЗАДАЧИ ОБ ОПТИМИЗАЦИИ ПАРАЛЛЕЛЬНОГО УПРАВЛЕНИЯ В ГАУССОВСКОЙ СЛУЧАЙНОЙ СРЕДЕ

А.В.Колногоров, А.Н.Лазутченко, Д.Н.Шиян

GENERALIZATION OF THE PROBLEM OF PARALLEL CONTROL OPTIMIZATION IN A GAUSSIAN RANDOM ENVIRONMENT

A.V.Kolnogorov, A.N.Lazutchenko, D.N.Shiiian

Институт электронных и информационных систем НовГУ, Alexander.Kolnogorov@novsu.ru

Рассматривается управление в гауссовской случайной среде, если для управления имеются два альтернативных действия, которым соответствуют доходы с априори известными дисперсиями и неизвестными математическими ожиданиями. Требуется определить наиболее эффективное действие и обеспечить его преимущественное применение. Задача допускает приложение к управлению обработкой больших объемов данных. Получено интегро-разностное уравнение, позволяющее определить оптимальное управление и величину максимальных потерь.

Ключевые слова: управление в случайной среде, задача о двуруком бандите, минимаксный и байесовский подходы, основная теорема теории игр, параллельная обработка

We consider control in a Gaussian random environment if there are two alternative actions corresponding to incomes with a priori known variances and unknown mathematical expectations. One should determine the most effective action and provide its preferable use. The results can be applied to control the processing of large amounts of data. We obtain an integro-differential equation for determination of the optimal control strategy and maximal losses.

Keywords: control in a random environment, two-armed bandit problem, minimax and Bayesian approaches, main theorem of the theory of games, parallel processing

1. Введение

Рассматривается задача об оптимизации параллельной многоэтапной обработки в случайной среде, продолжающая работы [1-3]. Пусть ξ_n , $n=1, \dots, N$ есть управляемый случайный процесс, значения которого интерпретируются как доходы, зависят только от выбираемых в текущие моменты времени действий y_n и имеют нормальные распределения с плотностями $f_{D_\ell}(x|m_\ell) = (2\pi D_\ell)^{-1/2} \exp(-(x-m_\ell)^2/(2D_\ell))$, если $y_n = \ell$ ($\ell=1,2$). Здесь D_1 , D_2 — известные дисперсии, а m_1 ,

m_2 — неизвестные математические ожидания дохода. Такая среда описывается векторным параметром $\theta = (m_1, m_2)$.

Цель управления состоит в максимизации (в некотором смысле) полного ожидаемого дохода. Для этого используется стратегия σ , которая в момент времени n является функцией текущей предыстории процесса, т. е. полученных откликов среды $x^{n-1} = x_1, \dots, x_{n-1}$ на выбранные действия $y^{n-1} = y_1, \dots, y_{n-1}$. Таким образом, $\sigma_\ell(y^{n-1}, x^{n-1}) = \Pr(y_n = \ell | y^{n-1}, x^{n-1})$, $\ell=1,2$. При $n=1$

предыстория отсутствует, поэтому в обозначениях она может опускаться. Множество стратегий обозначим Σ .

Если параметр θ известен, то следует всегда применять действие, которому соответствует большая из величин m_1 , m_2 и полный ожидаемый доход равен в этом случае $N(m_1 \vee m_2)$. Если же параметр неизвестен, то функция

$$L_N(\sigma, \theta) = E_{\sigma, \theta} \left(\sum_{n=1}^N ((m_1 \vee m_2) - \xi_n) \right)$$

характеризует потери дохода вследствие неполноты информации. Здесь $E_{\sigma, \theta}$ обозначает математическое ожидание по мере, порожденной стратегией σ и параметром θ . Множество допустимых значений параметра имеет вид: $\Theta = \{(m_1, m_2) : |m_1 - m_2| \leq 2C\}$, где C — некоторая константа ($0 < C < \infty$).

При минимаксном подходе максимальные потери на множестве параметров Θ минимизируются по множеству стратегий Σ , величина

$$R_N^M(\Theta) = \inf_{\Sigma} \sup_{\Theta} L_N(\sigma, \theta)$$

называется минимаксным риском, а соответствующая стратегия σ^M (если она существует) — минимаксной стратегией. Применение стратегии σ^M гарантирует выполнение неравенства $L_N(\sigma^M, \theta) \leq R_N^M(\Theta)$ при всех $\theta \in \Theta$, что означает робастность управления. Отметим, что другие подходы к робастному управлению в случайной среде рассматриваются, например, в [4-7], параллельное управление изучается в [8,9].

В данной работе используются результаты [1,2], где показано, что минимаксная стратегия и риск могут быть найдены как байесовские, соответствующие наихудшему априорному распределению. Эти результаты развиты на случай гауссовских случайных сред с различными дисперсиями. В разделе 2 получены рекуррентные уравнения для вычисления байесовского риска и потерь относительно наихудшего априорного распределения. В разделе 3 приведен пример нахождения минимаксных стратегий и риска численными методами для конкретной случайной среды. В разделе 4 дано заключение.

2. Стратегия параллельного управления

Как и в [1], положим $m_1 = u + v$, $m_2 = u - v$, тогда $\theta = (u + v, u - v)$, $\Theta = \{\theta : |v| \leq C\}$. В новых переменных асимптотически наихудшая плотность распределения имеет вид $v_a(u, v) = k_a(u) \rho(v)$, где $k_a(u)$ — постоянная плотность при $|u| \leq a$ и $a \rightarrow \infty$. Соответствующий байесовский риск равен

$$R_N^B(v_a(u, v)) = \inf_{\Sigma} \iint_{\Theta} L_N(\sigma, (u + v, u - v)) v_a(u, v) du dv. \quad (1)$$

Пусть к моменту времени $n = n_1 + n_2$ оба действия применены соответственно n_1 и n_2 раз, а X_1 , X_2 — полные доходы за их применение. Рассмотрим стратегию, которая в начале применяет каждое

действие по M_0 раз, а затем осуществляет оптимальное управление с тем ограничением, что действия могут меняться только после применения M раз подряд (считаем, что $2M_0 + kM = N$ для некоторого k). Отметим, что пакеты из M поступающих подряд данных, для обработки которых применяется одно и то же действие, могут обрабатываться параллельно. В этом случае полное время управления равно времени обработки $k+1$ пакетов данных и может быть сравнительно небольшим, если данные в пакете обрабатываются одновременно.

Дадим рекуррентное уравнение для вычисления байесовского риска, соответствующего стратегии параллельного управления с использованием вспомогательных рисков $R_{n_1, n_2}(Z)$, $n_1 \geq M_0$, $n_2 \geq M_0$.

Обозначим $n'_\ell = n_\ell / D_\ell$, $n^*_\ell = n_\ell \times D_\ell$, $M^*_\ell = M \times D_\ell$, $\ell = 1, 2$.

Теорема 1. Пусть $v_a(u, v) = k_a(u) \rho(v)$, где $k_a(u)$ — постоянная плотность при $|u| \leq a$ и $a \rightarrow \infty$. Для нахождения байесовского риска следует решить рекуррентное уравнение:

$$R_{n_1, n_2}(\cdot) = \min(R_{n_1, n_2}^{(1)}(\cdot), R_{n_1, n_2}^{(2)}(\cdot)),$$

где $R_{n_1, n_2}^{(1)}(Z) = R_{n_1, n_2}^{(2)}(Z) = 0$ при $n_1 + n_2 = N$,

$$R_{n_1, n_2}^{(1)}(Z) = M g_{n_1, n_2}^{(1)}(Z) + n_2^{-1} \int_{-\infty}^{+\infty} R_{n_1 + M, n_2}(Z + z) h_{n_1, n_2}^{(1)} \left(\frac{ZM - n_1 z}{n_2} \right) dz,$$

$$R_{n_1, n_2}^{(2)}(Z) = M g_{n_1, n_2}^{(2)}(Z) + n_1^{-1} \int_{-\infty}^{+\infty} R_{n_1, n_2 + M}(Z + z) h_{n_1, n_2}^{(2)} \left(\frac{ZM - n_2 z}{n_1} \right) dz$$

при $n_1 + n_2 < N$, $n_1 \geq M_0$, $n_2 \geq M_0$. Здесь

$$g_{n_1, n_2}^{(1)}(Z) = \int_{-C}^0 2|v| g_{n_1, n_2}(Z, v) \rho(v) dv,$$

$$g_{n_1, n_2}^{(2)}(Z) = \int_0^C 2v g_{n_1, n_2}(Z, v) \rho(v) dv,$$

$$g_{n_1, n_2}(Z, v) = (2\pi n_1^* n_2^* (n_1' + n_2'))^{-1/2} \exp \left(-\frac{(Z - 2vn_1n_2)^2}{2n_1^* n_2^* (n_1' + n_2')} \right),$$

$$h_{n_\ell, M}^{(\ell)}(z) = \left(\frac{n_\ell + M}{2\pi n_\ell M^*} \right)^{1/2} \exp \left(-\frac{z^2}{2n_\ell M^* (n_\ell + M)} \right).$$

Предельное значение байесовского риска (1) вычисляется по формуле

$$\begin{aligned} \lim_{a \rightarrow \infty} R_N^B(v_a(u, v)) &= R_N^B(\rho(v)) = \\ &= 2M_0 \int_{-C}^C v \rho(v) dv + \int_{-\infty}^{\infty} R_{M_0, M_0}(z) dz. \end{aligned} \quad (2)$$

Теорема 2. Для произвольной стратегии параллельного управления $\{\sigma_\ell(z, n_1, n_2)\}$ и плотности $v_a(u, v)$ вычисление функции потерь может быть выполнено следующим образом. Сначала следует решить рекуррентное уравнение

$$L_{n_1, n_2}(Z) = \sigma_1(Z, n_1, n_2) L_{n_1, n_2}^{(1)}(Z) + \sigma_2(Z, n_1, n_2) L_{n_1, n_2}^{(2)}(Z),$$

где $L_{n_1, n_2}^{(1)}(Z) = L_{n_1, n_2}^{(2)}(Z) = 0$ при $n_1 + n_2 = N$,

$$L_{n_1, n_2}^{(1)}(Z) = M g_{n_1, n_2}^{(1)}(Z) + n_2^{-1} \int_{-\infty}^{+\infty} L_{n_1, n_2, M}(Z+z) h_{n_1, M}^{(1)}\left(\frac{ZM-n_1 z}{n_2}\right) dz,$$

$$L_{n_1, n_2}^{(2)}(Z) = M g_{n_1, n_2}^{(2)}(Z) + n_1^{-1} \int_{-\infty}^{+\infty} L_{n_1, n_2, M}(Z+z) h_{n_2, M}^{(2)}\left(\frac{ZM-n_2 z}{n_1}\right) dz$$

при $n_1 + n_2 < N$, $n_1 \geq M_0$, $n_2 \geq M_0$. Тогда функция потерь вычисляется по формуле

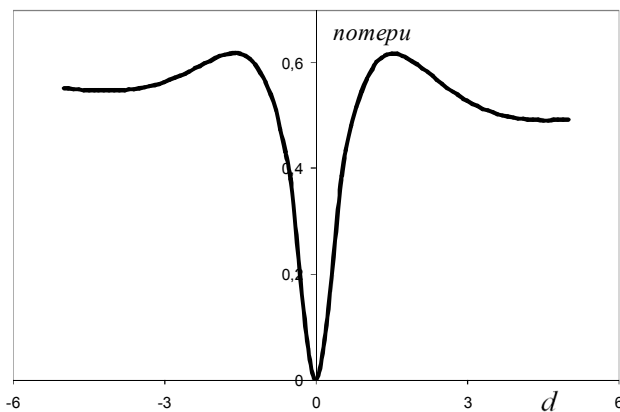
$$\lim_{a \rightarrow \infty} L_N^B(v_a(u, v)) = L_N^B(\rho(v)) = 2M_0 \int_{-C}^C v(\nu) d\nu + \int_{-\infty}^{\infty} L_{M_0, M_0}(z) dz. \quad (3)$$

Доказательства теорем аналогичны доказательствам, приведенным в [1, 2].

3. Численные эксперименты

Вычисления байесовского риска выполнялись по формуле (2) для среды с $D_1=1$, $D_2=0,75$ при $N=30$, $M_0=M=1$ в предположении, что $\rho(v)$ вырождена и сосредоточена в двух точках $v=-d_1 N^{-1/2}$ и $v=d_2 N^{-1/2}$ с вероятностями ρ_1 и ρ_2 соответственно. Вычисления выполнялись при $|v| \leq C = 5N^{-1/2} \approx 0,91$. Наихудшее априорное распределение соответствует максимуму нормализованного байесовского риска $N^{-1/2} R_N^B(\rho(v))$. Этот максимум был определен как приблизительно равный 0,62 при $d_1 \approx 1,60$, $d_1 \approx 1,54$, $\rho_1 \approx 0,524$.

Затем для найденной стратегии по формуле (3) были вычислены нормализованные потери $N^{-1/2} L_N^B(\rho(v))$. При $d < 0$ потери вычислялись для $\rho_1=1$, $d=-d_1$, при $d > 0$ — для $\rho_1=0$, $d=d_2$ соответственно. Соответствующая кривая представлена на рисунке. Два максимума кривой оказались приблизительно равны 0,62 при $d \approx -1,60$ и $d \approx 1,54$, что подтверждает сделанное предположение.



4. Заключение

Предложена стратегия параллельного управления в гауссовской случайной среде, характеризуемой различными дисперсиями доходов. Минимаксный риск ищется с помощью рекуррентного уравнения как байесовский, соответствующий наихудшему априорному распределению.

Работа выполнена при финансовой поддержке проектной части государственного задания в сфере

научной активности Министерства образования и науки Российской Федерации, проект №1.949.2014/К.

1. Kolnogorov A.V. Determination of the Minimax Risk for the Normal Two-Armed Bandit // Proc. of the IFAC Workshop ALCOSP'2010, Antalya, Turkey, August 26-28, 2010. <http://www.ifac-papersonline.net>
2. Колногоров А.В. Робастное параллельное управление в случайной среде (задача о двуруком бандите) // Автоматика и телемеханика. 2012. №4. С.114-130.
3. Лазутченко А.Н. О робастном управлении в случайной среде, характеризуемой нормальным распределением доходов с различными дисперсиями // Труды КарНЦ РАН. Сер.: Математическое моделирование и информационные технологии. 2015. №10. С.107-113.
4. Назин А.В., Позняк А.С. Адаптивный выбор вариантов. М.: Наука, 1986. 288 с.
5. Lugosi G., Cesa-Bianchi N. Prediction, learning and games., N. Y.: Cambridge University Press, 2006. 540 p.
6. Juditsky A., Nazin A.V., Tsybakov A.B., Vayatis N. Gap-free bounds for stochastic multi-armed bandit // Proc. 17th World Congress IFAC (Seoul, Korea, July 6-11, 2008). P.11560-11563.
7. Гасников А.В., Нестеров Ю.Е., Спокойный В.Г. Об эффективности одного метода рандомизации зеркального спуска в задачах онлайн оптимизации // Журн. вычисл. математики и математической физики. 2015. Т.55. №4. С.582-598.
8. Колногоров А.В., Мельникова С.В. Об оптимальной продолжительности начального этапа в двухэтапной модели целесообразного поведения в случайной среде // Вестник НовГУ. Сер.: Техн. науки. 2005. №34. С.73-75.
9. Lai T.L., Levin B., Robbins H., Siegmund D. Sequential Medical Trials (Stopping Rules/Asymptotic Optimality) // Proc. Natl. Acad. Sci. USA. 1980. V.77. №6. P.3135-3138.

References

1. Kolnogorov A.V. Determination of the Minimax Risk for the Normal Two-Armed Bandit. Proc. of the IFAC Workshop ALCOSP'2010. Antalya, Turkey, 2010. doi: 10.3182/20100826-3-TR-4015.00044. Available at: <http://www.ifac-papersonline.net>
2. Kolnogorov A.V. Robastnoe parallel'noe upravlenie v sluchainoi srede (zadache o dvurukom bandid) [Parallel design of robust control in the stochastic environment (the two-armed bandit problem)]. Avtomatika i Telemekhanika – Automation and Remote Control, 2012, vol. 73, no. 4, pp. 689-701.
3. Lazutchenko A.N. O robastnom upravlenii v sluchainoi srede, kharakterizuemoi normal'nym raspredeleniem dokhodov s razlichnymi dispersiiami [Robust control in a random environment characterized by a normal distribution of incomes with different variances]. Trudy KarNTs RAN. Ser. Matematicheskoe modelirovanie i informatsionnye tekhnologii – Trudy KarNTs RAN. Series: Mathematical Modeling and Information Technologies, 2015, no. 10, pp. 107-113.
4. Nazin A.V., Pozniak A.S. Adaptivnyi vybor variantov [Adaptive choice of alternatives]. Moscow, "Nauka" Publ., 1986. 288 p.
5. Lugosi G., Cesa-Bianchi N. Prediction, learning and games. Cambridge University Press, New York, 2006. 540 p.
6. Juditsky A., Nazin A. V., Tsybakov A.B., Vayatis N. Gap-free bounds for stochastic multi-armed bandit. Proc. of the 17th World Congr. of the International Federation of Automatic Control (IFAC 2008). Seoul, Korea, 2008, vol. 17, pp. 11560-11563.
7. Gasnikov A. V., Nesterov Iu. E., Spokoinyi V. G. Ob effektivnosti odnogo metoda randomizatsii zerkal'nogo spуска v zadachakh onlain optimizatsii [On the efficiency of a randomized mirror descent algorithm in online optimization problems]. Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki – Computational Mathematics and Mathematical Physics, 2015, vol. 55, no. 4, pp. 580-596.
8. Kolnogorov A.V., Mel'nikova S.V. Ob optimal'noi prodolzhitel'nosti nachal'nogo etapa v dvukhetapnoi modeli tselesoobraznogo povedeniia v sluchainoi srede [On the optimal duration of the first stage in two-stage model of expedient behavior in random medium]. Vestnik NovGU. Ser. Tekhnicheskie nauki – Vestnik NovSU. Issue: Engineering Sciences, 2005, no. 34, pp. 73-75.
9. Lai T.L., Levin B., Robbins H., Siegmund D. Sequential Medical Trials (Stopping Rules/Asymptotic Optimality). Proc. of the National Academy of Sciences, 1980, v.77, no.6, p.3135-3138.