Асимптотический метод вычисления многомерных интегралов в задачах прогнозирования состояния термокарстовых озер*

Ю. С. Попков[I], Ю. М. Полищук[II]

[I] Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия
[II] Югорский НИИ информационных технологий, Ханты-Мансийск, Россия

Аннотация. Статья посвящена разработке аналитического метода приближенного вычисления многомерных интегралов, ориентированного на решение балансовых уравнений в процедурах рандомизированного машинного обучения. Последние используются для прогнозирования эволюции площади термокарстовых озер. Метод базируется на разложении в ряд аналитической функции - экспоненты - и трансформации многомерных интегралов в произведение простых одномерных интегралов на интервальных множествах.

Ключевые слова: термокарстовые озера; ряд Тейлора; рандомизированное машинное обучение; метод Лагранжа; балансовые уравнения.

DOI 10.14357/20718632240208 **EDN** WWJQDJ

Введение

Проблемы изменения климата привлекают внимание научного сообщества и общественности, о чем свидетельствуют многочисленные публикации в специализированных изданиях и средствах массовой информации. В качестве популярной причины этих изменений декларируются так называемые парниковые газы, генерируемые как человеческой деятельностью, так и природными процессами. К последним относятся процессы генерации метана термокарстовыми озерами в арктических зонах земной поверхности [1, 3], в частности, в арктической зоне РФ [2, 4]. Исследованию этих процессов посвящена работа [5, 6], в которой развивался метод рандомизированного машинного обучения (РМО), промежуточный результат которого сводился к необходимости вычисления многомерных интегралов на простых множествах (многомерных параллелепипедах). При разумной размерности исходной задачи эти вычисления сопряжены с большими и до конца не разрешенными сложностями. Последние связаны с генерацией подходящих случайных последовательностей, технологией вычисления многомерных интегралов в рамках процедуры Монте Карло [7].

В данной статье предлагается асимптотический метод аналитического вычисления многомерных интегралов, ориентированный на класс многомерных интегралов, присутствующих в балансовых уравнениях процедуры РМО, которая используется для обучения моделей и прогнозирования эволюции термокарстовых озер.

* Работа выполнена при поддержке гранта РНФ (проект 22-11-20023).

1. Математические основы процедуры РМО

Ядром процедуры РМО является математическая модель формирования термокарстовых озер, устанавливающая связь площади озер $S[n]$ с климатическими параметрами, а именно, средними температурой $T[n]$ и осадками $R[n]$ (n - момент измерения, обычно год).

Связь между указанными переменными описывается математической моделью *линейной динамической регрессии* с памятью p:

$$S[n] = a_0 + \sum_{k=1}^{p} a_k\, S[n-k] + a_{p+1}\, T[n] + a_{p+2}\, R[n]. \tag{1}$$

Пусть имеются данные по площади $S^r[n]$ и климатическим параметрам $T^r[n], R^r[n]$ на интервале наблюдений на $\mathcal{T} = [p, p+N]$. Обозначим выход модели в момент времени n, когда в правую часть равенства (1) входят реальные данные, в виде

$$S[n|a] = a_0 + \sum_{k=1}^{p} a_k\, S^r[n-k] + a_{p+1}\, T^r[n] + a_{p+2}\, R^r[n], \tag{2}$$

$$S[n|a] = \langle s_n^r, a\rangle = \sum_{i=0}^{p+2} a_i s_{ni}^r, \tag{3}$$

$$n = \overline{p, p+N},$$

где:

- векторы наблюдаемых значений площади, температуры и осадков в момент n

$$s_n^r = \{1, S^r[n-1], \dots, S^r[n-p], T^r[n], R^r[n]\},\ n = \overline{p, p+N}; \tag{4}$$

- s_{ni}^r - компоненты вектора (4);
- вектор параметров модели

$$a = \{a_0, \dots, a_p, a_{(p+1)}, a_{(p+2)}\}; \tag{5}$$

Параметры модели a - случайные и интервальные:

$$a \in \mathcal{A} = [a^-, a^+],\ \mathcal{A} = \bigcup_{i=0}^{p+N} \mathcal{A}_i,\ \mathcal{A}_i = [a_i^-, a_i^+]. \tag{6}$$

Функции плотности распределения вероятностей (ПРВ) параметров $P(a)$ предполагаются непрерывно-дифференцируемыми функциями.

Тогда, согласно концепции РМО [5] энтропийно-оптимальная ПРВ параметров имеет следующий вид:

$$P^*(a) = \frac{\exp\left(-\sum_{n=p}^{p+N} \lambda_n\, S[n|a]\right)}{\mathcal{P}(\boldsymbol{\lambda})}, \tag{7}$$

$$\mathcal{P}(\boldsymbol{\lambda}) = \int_{\mathcal{A}} \exp\left(-\sum_{n=p}^{p+N} \lambda_n\, S[n|a]\right)\, da, \tag{8}$$

Множители Лагранжа $\boldsymbol{\lambda} = \{\lambda_p, \dots, \lambda_{p+N}\}$ определяются решением следующей системы уравнений, которые содержат многомерные интегралы:

$$\int_{\mathcal{A}} \exp\left(-\sum_{n=p}^{p+N} \lambda_n\, S[n \mid a]\right) (S[n \mid a] - S^r[n])\, da = 0, \tag{9}$$

$$n = \overline{p, p+N}.$$

Согласно (6) множествами определения интегралов являются многомерные параллелепипеды.

2. Асимптотический метод вычисления многомерных интегралов

Из равенств (9) следует, что подынтегральные функции содержат аналитическую функцию - экспоненту, которая апроксимируется степенным рядом (Тейлора-Маклорена), сходящимся на всей числовой оси (т.е. при всех значениях аргумента этой функции).

Обозначим показатель степени в экспоненте x и поменяем порядок суммирования:

$$x = \sum_{n=p}^{p+N} \lambda_n\, S[n|a] = \sum_{i=0}^{p+2} a_i B_i^r(\boldsymbol{\lambda}), \tag{10}$$

где

$$S[n|a] = \sum_{i=0}^{p+2} a_i s_{ni}^r \tag{11}$$

$$B_i^r(\boldsymbol{\lambda}) = \left(\sum_{n=p}^{(p+N)} \lambda_n s_{ni}^r\right)$$

$$x_n \in (-\infty, +\infty).$$

Воспользуемся аналитичностью экспоненты и представим ее следующим полиномом Тейлора-Маклорена степени m вокруг $x_0 = 0$:

$$\exp(-x) = 1 - x + \frac{1}{2}\, x^2 + \cdots + (-1)^m \frac{1}{m!}\, x^m + \ldots \tag{12}$$

Согласно (10, 12) получим следующий степенной ряд:

$$\exp(-x) = 1 - \left(\sum_{i=0}^{p+2} a_i B_i^r(\boldsymbol{\lambda})\right) + \frac{1}{2}\left(\sum_{i=0}^{p+2} a_i B_i^r(\boldsymbol{\lambda})\right)^2 + \ldots \tag{13}$$

Подставляя это равенство в (9), получим N-мерный полином относительно компонент вектора множителей Лагранжа $\boldsymbol{\lambda}$, в коэффициенты которого входят многомерные интегралы относительно параметров a линейной модели динамической регрессии, характеризующей временную эволюцию термокарстовых озер.

Многомерные интегралы, входящие в эти коэффициенты, имеют следующий общий вид:

$$I_{m_0,\ldots,m_k}(i_0, \ldots, i_k;\ q_0, \ldots, q_{p+3-k}) = \underbrace{\int \cdots \int}_{\mathcal{A}_k} a_{i_0}^{m_0} \cdots a_{i_k}^{m_k}\, da_{i_0} \cdots da_{i_k} \times$$
$$\times \underbrace{\int \cdots \int}_{\mathcal{A}_{l=p+3-k}} da_{q_0} \cdots da_{q_l}, \tag{14}$$
$$(i_0, \ldots, i_k) \neq (q_0, \ldots, q_l),$$
$$(i_\mu, q_\nu) = \overline{0, p+2}, \mu = \overline{0, k}, \nu = \overline{0, l},\ ,\ k + l = p + 3.$$

где:

$$m_j = 1,2, \ldots, m,\ j = \overline{1, k};\ m_0 + \cdots + m_k \leq m, \tag{15}$$

m - степень полинома в (12).

Первая группа из k интегралов

$$I_{(m_0,\ldots,m_k)} = \underbrace{\int \cdots \int}_{\mathcal{A}_k} a_{i_0}^{m_0} \cdots a_{i_k}^{m_k}\, da_{i_0} \cdots da_{i_k} = \prod_{j=0}^{k} \frac{(a_{i_j}^{+})^{(m_j+1)} - (a_{i_j}^{-})^{(m_j+1)}}{m_j + 1}. \tag{16}$$

Вторая группа из $p + 3 - k$ интегралов

$$I_{(q_0,\ldots,q_{(p+3-k)})} = \underbrace{\int \cdots \int}_{\mathcal{A}_{l=p+3-k}} da_{q_0} \cdots da_{q_l} = \prod_{j=0}^{p+3-k} \left(a_{q_j}^{+} - a_{q_j}^{-}\right). \tag{17}$$

Если размеры множеств значений параметров одинаковые для всех параметров и равные $[b, a]$, то

$$I_{(m_0,\ldots,m_k)} = \prod_{j=0}^{k} \frac{b^{(m_j+1)} - a^{(m_j+1)}}{m_j + 1}. \tag{18}$$

Таким образом, аппроксимация степенным рядом аналитической функции и параметрическая линейность модели позволяют трансформировать многомерные интегралы к произведениям одномерных интегралов на интервальных множествах. Подынтегральные функции в этих интегралах представляют собой степени параметров.

3. Пример

Для иллюстрации предложенного выше асимптотического метода вычисления многомерных интегралов рассмотрим задачу рандомизированного обучения линейной модели эволюции площади термокарстовых озер с памятью $p = 1$ и параметром $a_3 = 0$ [1].

$$S[n] = a_0 + a_1\, S[n-1] + a_2\, T[n]. \tag{19}$$

Интервал обучения $\mathcal{T}_l = [1,2]$, $(p = 1, N = 1)$. Поскольку модель имеет память $p = 1$, то согласно (19) необходима информация о значении площади $S^r[0]$. Таким образом, выходы модели в интервале обучения можно представить в виде:

[1] В предположении, что влияние осадков существенно меньше влияния температуры.

$$S[1|a] = a_0 + a_1\, S^r[0] + a_2\, T^r[1] = \sum_{i=0}^{2} a_i s_{1i}^r$$
$$S[2|a] = a_0 + a_1\, S^r[1] + a_2\, T^r[2] = \sum_{i=0}^{2} a_i s_{2i}^r, \qquad (20)$$

где

$$s_{10}^r = 1,\ s_{11}^r = S^r[0],\ s_{12}^r = T^r[1],$$
$$s_{20}^r = 1,\ s_{21}^r = S^r[1],\ s_{22}^r = T^r[2]. \qquad (21)$$

Переменные B из (11) в данном примере имеют вид:

$$B_0(\boldsymbol{\lambda}) = \lambda_1 + \lambda_2,\ B_1(\boldsymbol{\lambda}) = \lambda_1\, S^r[0] + \lambda_2\, S^r[1],$$
$$B_2(\boldsymbol{\lambda}) = \lambda_1\, T^r[1] + \lambda_2\, T^r[2]. \qquad (22)$$

Заметим, что они являются линейными функциями множителей Лагранжа. Интервалы по параметрам одинаковые и равные [0,1].

Напомним, что верхний индекс в этих равенствах означает, что соответствующие величины - реальные данные.

В результате применения РМО-процедуры получаем следующие два балансовых уравнения для определения множителей Лагранжа:

$$\int_0^1 \int_0^1 \int_0^1 \exp(-\lambda_1 S[1|a] + \lambda_2\, S[2|a])(S[1|a] - S^r[1])da = 0;$$
$$\int_0^1 \int_0^1 \int_0^1 \exp(-\lambda_1 S[1|a] + \lambda_2\, S[2|a])(S[2|a] - S^r[2])da = 0,$$

Воспользуемся представлением экспоненты полиномом 2-го порядка и правилами вычисления многомерных интегралов с полиномиальными подынтегральными функциями (18). Получим систему из двух балансовых уравнений с полиномами второго порядка по множителям Лагранжа следующего вида:

$$\frac{1}{2}S^r[1]\sum_{i=0}^{2} B_i(\boldsymbol{\lambda}) - \frac{1}{4}\sum_{(i_1,i_2)=0}^{2} B_{(i_1)}(\boldsymbol{\lambda})\, s_{(1,i_1)}^r - \frac{1}{8}S^r[1]\sum_{(i_1,i_2)=0}^{2} B_{(i_1)}(\boldsymbol{\lambda})\, s_{(1,i_2)}^r -$$
$$-\frac{1}{8}S^r[1]\sum_{(i_1,i_2)=0}^{2} B_{(i_1)}(\boldsymbol{\lambda})\, B_{(i_2)}(\boldsymbol{\lambda}) + \frac{1}{8}\sum_{(i_1,i_2,i_3)=0}^{2} B_{(i_1)}(\boldsymbol{\lambda})\, B_{(i_2)}(\boldsymbol{\lambda})\, s_{(1,i_3)}^r = \frac{1}{2}\sum_{i=0}^{2} s_{(1,i)}^r - S^r[1],$$

$$\frac{1}{2}S^r[2]\sum_{i=0}^{2} B_i(\boldsymbol{\lambda}) - \frac{1}{4}\sum_{(i_1,i_2)=0}^{2} B_{(i_1)}(\boldsymbol{\lambda})\, s_{(2,i_1)}^r - \frac{1}{8}S^r[2]\sum_{(i_1,i_2)=0}^{2} B_{i_1}(\boldsymbol{\lambda})\, s_{(2,i_2)}^r -$$
$$-\frac{1}{8}S^r[2]\sum_{(i_1,i_2)=0}^{2} B_{(i_1)}(\boldsymbol{\lambda})\, B_{(i_2)}(\boldsymbol{\lambda}) + \frac{1}{8}\sum_{(i_1,i_2,i_3)=0}^{2} B_{(i_1)}(\boldsymbol{\lambda})\, B_{(i_2)}(\boldsymbol{\lambda})\, s_{(2,i_3)}^r = \frac{1}{2}\sum_{i=0}^{2} s_{(2,i)}^r - S^r[2],$$

Таким образом, балансовые уравнения трансформируются в «приближенно-эквивалентные» уравнения с полиномами второй степени в правой части. Для такого класса уравнений удается сформировать рекуррентную процедуру для поиска аналитического решения. Последнему будут посвящены направления дальнейших исследований.

4. Обсуждение

Обратимся к балансовым уравнениям (9), в которые входят многомерные интегралы от экспоненциальных функций, показатели которых представляют собой линейные функции параметров. Не трудно видеть, что подобные интегральные конструкции поддаются аналитичекому вычислению. Однако получаемые при этом выражения приводят к балансовым уравнениям относительно множителей Лагранжа, состояшим из дробно-линейных функций по множителям Лагранжа. Решение такого класса уравнений удается провести только численно, и при том, что априорные свойства их установить невозможно. Последнее делает непредсказуемым результат применяемых вычислительных процедур.

Поэтому применение предлагаемого асимптотического метода, приводящего к аналитически вычисляемым многомерным интегралом, является принципиально и вычислительно существенно эффективнее. К тому же, следует отметить, что в результате применения асимптотического метода приближенные балансовые уравнения имеют полиномиальную левую часть, что позволяет адаптировать метод абстрактных степенных рядов с целью формирования аналитической рекуррентной процедуры для их решения.

Заключение

В статье развивается асимптотический аналитический метод приближенного вычисления многомерных интегралов, ориентированный на задачи прогнозирования эволюции площади термокарстовых озер с помощью процедуры рандомизированного машинного обучения. Используется разложение экспоненциальной функции в степенной ряд и развито преобразование многомерных интегралов к произведению одномерных интегралов над степенными функциями.

Литература

1. Zuidhoff F.S., Kolstrup E. Changes in palsa distribution in relation to climate change in Laivadalen, Northern Sweden, espesially 1960-1997. Permafrost and Periglacial Processes, 2000, v.11, pp. 55-69.
2. Kirpotin S., Polishchuk Y., Bruksina N. Abrupt changes of thermokarst lakes in Western Siberia: impacts of climatic warming on permafrost melting. International Journal of Environmental Studies. 2009, v. 66, No.4, pp.423-431.
3. Karlson J.M., Lyon S.W., Destouni G. Temporal behavior of lake size-distribution in a thawing permafrost landscape in Northwestern Siberia. Remote sensing, 2014, No. 6, pp. 621-636.
4. Bryksina N.A., Polishchuk Yu.M. Analysis of changes in the number of thermokarst lakes in permafrost of Western Siberia on the basis of satellite images. Cryosphere of Earth, 2015, v. 19, No.2, pp. 114-120.
5. Popkov Y.S., Popkov A.Y., Dubnov Y.A. Entropy randomization in Machine Learning. CRC Press, 2023
6. Dubnov Y.A., Popkov A.Y., Polishcuk V.Y., Sokol E.S., Melnikov A.V., Y.M.Polishcuk Y.M., Popkov Y.S. Randomized Mashine Learning Algoritms to Forecast the Evolution of Thermokarst Lakes Area in Permafrost Zones. Automation and Remote Control, 2023, v.84, No.1, p. 56-70.
7. Дарховский Б.С., Попков А.Ю., Попков Ю.С. Метод пакетных итераций Монте-Карло для решения задач глобальной оптимизации. Информационные технологии и вычислительные системы. 2014, вып.3. С. 39-52.

Попков Юрий Соломонович. Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия. Главный научный сотрудник; Академик РАН, доктор технических наук, профессор. Область научных интересов: энтропийные методы, макросистемы, рандомизированное машинное обучение. E-mail: popkov@isa.ru

Полищук Юрий Михайлович. Югорский НИИ информационных технологий, г. Ханты- Мансийск, Россия. Главный научный сотрудник, доктор физико-математических наук. Область научных интересов: информационные технологии. E-mail: yupolishchuk@gmail.com

Assimptotic Numerical Method for Multidimensional Integrals of Forecasting of Thermokarst Lakes

Y. S. Popkov[I], Y. M. Polyschuk[II]

[I] Federal Research Center “Computer Science and Control” of RAS, Moscow, Russia
[II] Yugra Research Institute of Information Technologies, Khanty-Mansyisk, Russia

Abstract. We develop an analytical method for the approximate calculation of multidimensional integrals, focused on solving balance equations in Randomized Machine Learning procedures. The latter are used to forecast the evolution of thermokarst lakes’ area. The method is based on the series expansion of an analytical function - the exponential - and the transformation of multidimensional integrals into the product of simple one-dimensional integrals on interval sets.

Keywords: thermokarst lakes; Taylor series; randomized machine learning; Lagrange method; balance equations.

DOI 10.14357/20718632240208 **EDN** WWJQDJ

References

1. Zuidhoff F.S., Kolstrup E. Changes in palsa distribution in relation to climate change in Laivadalen, Northern Sweden, espesially 1960-1997. Permafrost and Periglacial Processes, 2000, v.11, pp. 55-69.
2. Kirpotin S., Polishchuk Y., Bruksina N. Abrupt changes of thermokarst lakes in Western Siberia: impacts of climatic warming on permafrost melting. International Journal of Environmental Studies. 2009, v. 66, No.4, pp.423-431.
3. Karlson J.M., Lyon S.W., Destouni G. Temporal behavior of lake size-distribution in a thawing permafrost landscape in Northwestern Siberia. Remote sensing, 2014, No. 6, pp. 621-636.
4. Bryksina N.A., Polishchuk Yu.M. Analysis of changes in the number of thermokarst lakes in permafrost of Western Siberia on the basis of satellite images. Cryosphere of Earth, 2015, v. 19, No.2, pp. 114-120.
5. Popkov Y.S., Popkov A.Y., Dubnov Y.A. Entropy randomization in Machine Learning. CRC Press, 2023
6. Dubnov Y.A., Popkov A.Y., Polishcuk V.Y., Sokol E.S., Melnikov A.V., Y.M.Polishcuk Y.M., Popkov Y.S. Randomized Mashine Learning Algoritms to Forecast the Evolution of Thermokarst Lakes Area in Permafrost Zones. Automation and Remote Control, 2023, v.84, No.1, p. 56-70.
7. Darkhovsky B.S., Popkov A.Y., Popkov Y.S. Method of Batch Monte Carlo Iterations for Solving of Global Optimization Problems // Informacionnie Technologii i Vichislitelnie Sistemy, 2014, No.3. pp. 39-52.

Popkov Yuri S. Federal Reerach Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia, chief research scientist, academician of Russian Academy of Sciences, doctor of science, professor. Scientific area: entropy methods, macrosystems, randomized machine learning. E-mail: popkov@isa.ru..

Polyschuk Yuri M. Yugra Research Institute of Information Technologies, Khanty-Mansyisk, Russia, chief research scientist, doctor of science. Scientific area: information technologies. E-mail: yupolishchuk@gmail.com