

АЛГОРИТМ ОЦІНЮВАННЯ НЕВІДОМОГО ПАРАМЕТРА ГІББСОВСЬКОГО РОЗПОДІЛУ НА ОСНОВІ МЕТОДУ СТОХАСТИЧНИХ КВАЗІГРАДІЄНТІВ¹

Анотація. Розглянуто практичний алгоритм оцінювання невідомого параметра математичної моделі марковського процесу з локальною взаємодією на основі гіббсовського розподілу. Запропоновано застосувати метод стохастичних квазіградієнтів до функції максимальної вірогідності, що побудована за спостереженнями реалізацій гіббсовського поля. Отримано результати, які мають широку прикладну сферу застосування для моделювання стохастичних процесів.

Ключові слова: гіббсовський розподіл, метод максимальної вірогідності, марковські випадкові поля, стохастичний квазіградієнтний метод, оцінювання параметрів.

Задача синтезу нових текстур на основі певного набору зображень полягає у створенні нових конфігурацій, візуально схожих на вхідну модель. Саме тому в алгоритмах, що застосовують в цій предметній області, поєднуються результати багатьох споріднених досліджень, таких, наприклад, як теорія розпізнавання образів, методи опуклої оптимізації, інструментарій випадкових полів тощо [1]. Одним із поширених підходів для розв'язання таких задач є використання стохастичних моделей, наприклад гіббсовських полів, де значення параметрів кожного пікселя залежать від значень сусідніх пікселів. У цьому випадку для генерації нових зразків потрібно знайти параметри текстури, які і характеризують силу взаємодії сусідніх пікселів.

Побудова моделей зображення на основі гіббсовських випадкових полів базується на певній геометричній структурі і кількісних характеристиках взаємодії пікселів. Стохастична текстура таких моделей складається з графу сусідства, де пікселі є вершинами графу V , а ребра E сполучають пари пікселів, що взаємодіють між собою. Носієм взаємодії пікселів є набір інваріантних до зсуву повних підграфів графу сусідства — клік. Кожна кліка має власний потенціал, який залежить від сигналів (станів елементів) у піксельній парі. Взаємодія пікселів означає, що однакові яскравості в кожній окремій кліці не можуть набувати абсолютно випадкових значень.

Інакше кажучи, кожній кліці χ ставиться у відповідність потенціал-функція $H(x_\chi, v)$, яка залежить від параметрів, згенерованих реалізацією поля X у точках кліки, і деякого невідомого параметра $v \in \Theta$ (Θ — опукла множина). Тут x_χ — конфігурація кліки χ , тобто $x_\chi = \{x_{i_1}, x_{i_2}, \dots, x_{i_{|\chi|}}\}$, де $i_j \in \chi$. Функції $H(x_\chi, v)$ задовольняють умову $|H(x_\chi, v)| < \infty$. Будемо вважати, що потенціали задані з точністю до деяких параметрів, які складають параметричний вектор, а розподіл імовірностей реалізації певної конфігурації поля можна записати як гіббсовський розподіл

$$P(X = x) = Z(v)^{-1} \exp \left(- \sum_{\chi \in \mathbb{N}} H(x_\chi, v) \right), \quad (1)$$

де $Z(v)$ — нормувальний множник, сумування здійснюється по всій множині

¹ Робота виконана за часткової підтримки Національного фонду досліджень України. Грант № 2020.02/0121.

© О.С. Самосьонук, 2023

конфігурацій елементів поля

$$Z(v) = \sum_{x \in \Lambda} \exp \left(- \sum_{\chi \in \mathbb{N}} H(x_\chi, v) \right).$$

Додамо до моделі умову марковості, а саме будемо вважати, що стан деякого елемента поля залежить лише від сусідніх елементів поля. Послідовність станів такої системи локально незалежна в сенсі

$$\prod_{s \in T} P(x_s | x_{S \setminus s}) = \prod_{s \in T} P(x_s | x_{\partial s}),$$

де T — деяка підмножина множини вершин поля S , ∂s — окіл вершини s , що розглядається.

Вигляд умовних імовірностей станів одного елемента поля залежить від структури зв'язків між елементами поля, але узгоджений набір умовних імовірностей (тобто такий, що гарантує існування спільного розподілу) дає змогу повністю описати випадкове поле X . Запишемо їхній загальний вигляд як

$$P(x_s | x_{\partial s}) = Z_s(v)^{-1} \exp(-U_s(x, v)), \quad (2)$$

де $U_s(x, v)$ — сума потенціалів всіх клік, до яких належить елемент s , так звана енергетична функція по околу ∂s елемента s

$$U_s(x, v) = \sum_{\chi \supset s} H(x_\chi, v),$$

$Z_s(v)$ — нормувальний множник

$$Z_s(v) = \sum_{x_s \in A} \exp(-U_s(x, v)).$$

На відміну від спільного розподілу (1), де міститься нормувальна константа, яку практично неможливо обчислити (для цього потрібне сумування по всьому простору конфігурацій елементів поля $|\Lambda| = |A|^{|S|}$), умовні розподіли (2) зазвичай є простішими для обчислення, оскільки містять суму лише по станах одного елемента поля.

Таким чином, маючи послідовність незалежних спостережень $\{x^1, \dots, x^n\}$, вважаємо, що $x_s^i \in A$, $i = 1, 2, \dots, n$, позначає стан деякого елемента s в i -му спостереженні. А саме процес пошуку оптимальної структури зображення полягає в пошуку істинного значення невідомого параметра $v^* \in \Theta$ на основі результатів спостережень $x \in \Lambda$ і відомої енергії $U_s(x, v)$.

Після деталізації і опису математичної моделі наступним етапом є безпосереднє розв'язання задачі оцінювання невідомого параметра v^* .

Поширеним підходом є використання алгоритму методу максимальної вірогідності (Maximum Likelihood Estimation, MLE) або Баєсових процедур розпізнавання. Метод MLE дає змогу знайти параметр, за якого імовірність спостереження отриманих даних максимальна, тоді як Баєсові процедури використовують оцінку апостеріорного максимуму, що дає змогу обчислити деяку величину з використанням апіорного розподілу ймовірностей. Далі розглянемо метод максимальної вірогідності.

Для оцінювання невідомих параметрів гіббсовського розподілу (2) використовують функцію правдоподібності $L_n(x, v)$, яка будується за послідовністю $\{x^1, x^2, \dots, x^n\}$ незалежних спостережень реалізацій випадкового поля X — так званої навчальної вибірки у такий спосіб:

$$L_n(x, v) = P(x^1, v) \cdot P(x^2, v) \cdot \dots \cdot P(x^n, v).$$

Так, наприклад, для широковідомої моделі Ізінга функція максимальної правдоподібності може бути записана як добуток імовірностей, що кожен піксель у малюнку набуває певного значення

$$L_n(x, v) = \prod_i \prod_s P(x_s^i = 1 / x_{cs}^i).$$

Тоді оцінкою оптимального параметра v^* буде розв'язок задачі максимізації функції максимальної правдоподібності $v_n = \arg \max L_n(x, v)$.

Запишемо від'ємну функцію логарифмічної правдоподібності, врахувавши умовний розподіл (2)

$$L_n(x, v) = \sum_{i=1}^n (U(x^i, v) + \ln(Z(x^i, v))),$$

і розглянемо задачу пошуку її мінімуму $v_n = \arg \min L_n(x, v)$.

Безпосереднє застосування методу максимальної вірогідності для знаходження аналітичного виразу оцінки невідомого параметра ускладнено через те, що нормуючий коефіцієнт залежить від параметра v^* [2]. Тому для розв'язання цієї задачі розглянемо стохастичний квазіградієнтний метод, запропонований у [3].

Метод стохастичних квазіградієнтів є різновидом процедур стохастичної оптимізації, що враховує додаткову інформацію про кривизну цільової функції. У звичайному методі стохастичних градієнтів спуск розраховується виключно на основі деякої обмеженої підмножини доданків цільової функції, які можуть містити шум або неточно відображати її загальну структуру. Метод стохастичних квазіградієнтів використовується для усунення цих обмежень, оскільки враховує інформацію про попередні обчислення градієнтів для наближення до справжнього градієнта цільової функції. Інакше кажучи, використовується квазіградієнт, який поєднує поточний градієнт, обчислений за вибіркою з доданків цільової функції, з розрахунками градієнтів, отриманими на попередніх кроках. Ідея полягає в тому, щоб використовувати попередні оцінки градієнта для отримання кращого наближення, що дасть потенційно точніші оновлення параметрів моделі. В загальному вигляді для мінімізації опуклої функції

$$F(v) = E_x \{f(v, x)\} \rightarrow \min_{v \in \Theta}$$

будується рекурентна послідовність

$$v_{k+1} = \pi_{\Theta}(v_k - \rho_k \gamma^k), \quad k = 0, 1, \dots \quad (3)$$

Тут v_0 — довільна початкова точка; v_k — точка, отримана після k -го кроку ітерації; π_{Θ} — оператор проєктування на опуклу множину Θ ; ρ_k — величина кроку спуску, $\rho_k \geq 0$; γ^k — стохастичний напрямок спуску для функції $F(v)$. У загальному випадку напрямки γ^k є оцінками градієнтів у точці x^i та мають задовольняти умову

$$E(\gamma^k | v_0, \dots, v_k) = a_k \hat{F}_v(v_k) + b^k, \quad k = 0, 1, \dots, \quad (4)$$

де a_k — невід'ємна випадкова величина та $b^k = (b_1^k, \dots, b_n^k)$ — випадковий вектор помилок залежать від послідовності $(v_0, \dots, v_k)$. Вектор $\hat{F}_v(v_k)$ — узагальнений градієнт, тобто вектор, що задовольняє для будь-яких z нерівність $F(z) - F(v_k) \geq (\hat{F}_v(v_k), z - v_k)$. При цьому випадковий вектор γ^k називається стохастичним квазіградієнтом функції $F(v)$ у точці v_k для $b^k \neq 0$.

У розгляданій задачі потрібно мінімізувати функцію

$$v_n = \arg \min_{v \in \Theta} L_n(x, v) = \arg \min_{v \in \Theta} \sum_{i=1}^n f^i(v),$$

де $f^i(v) = U(x^i, v) + \ln(Z(v))$ — опуклі вниз неперервно диференційовані функції за умови $v \in \Theta$, множина Θ опукла і замкнута.

Отже, алгоритм пошуку мінімуму функціонала $L_n(x, v)$ такий: довільно вибираємо початкову точку v_0 , знаходимо значення цільової функції $F_n(x, v_0)$ і, поклавши $k=1$, переходимо до основної частини алгоритму.

1. З імовірністю $\frac{1}{n}$ для $v = v_k$ вибираємо одне з чисел $i=1, \dots, n$. Позначаємо вибране число i_{fix} . Для $\nabla f^{i_{fix}}(v_k)$ справедливо

$$E(\nabla f^{i_{fix}}(v_k) | v_k) = \frac{1}{n} \sum_{i=1}^n \nabla f^i(v_k) = \frac{1}{n} \nabla F(x, v_k).$$

Це означає, що оскільки функція $F_n(v)$ складається з суми n блоків $f^1(v), \dots, f^n(v)$, то в k -й ітерації можемо вибрати з однаковою ймовірністю один з доданків $f^i(v_k)$ і як напрямок спуску вибрати антиградієнт, розрахований тільки для цього блока.

2. Обчислюємо стохастичний квазіградієнт γ^k у точці v_k . Оскільки функції $f^i(v)$ неперервно диференційовані, то

$$\gamma^k = a_k \left(\frac{\partial f^{i_{fix}}(v_k)}{\partial v} \right) = a_k \left(\frac{\partial U(x^i, v_k)}{\partial v} + \frac{1}{Z(v)} \frac{\partial Z(v)}{\partial v} \right), \quad a_k \geq 0.$$

Значення a_k вибираються з урахуванням умови $\|\gamma^k\|^2 \leq C$, C — const, і здебільшого залежать від конкретного вигляду потенціальних функцій.

3. Вибираємо величину крокового множника ρ_k , враховуючи умови $\sum_{k=0}^{\infty} \rho_k^2 < \infty$, $\sum_{k=0}^{\infty} \rho_k a_k = \infty$, необхідні для збіжності алгоритму. В загальному вигляді проблема вибору величини кроку не є тривіальною. Якщо крок вибирається малим (щоб гарантувати збіжність), то процедура збігається повільно, у разі збільшення кроку (для пришвидшення збіжності) алгоритм може перескочити точку мінімуму. В подальших обчисленнях пропонується використовувати значення $\rho_k = \frac{1}{k}$. Детальнішу інформацію про способи вибору крокового множника і відповідні швидкості збіжності наведено в [3, 4].

4. Знаходимо точку $v_{k+1} = \pi_{\Theta} \left(v_k - \frac{1}{k} \gamma^k \right)$. Розглянемо детальніше оператор проєктування вектора $\left(v_k - \frac{1}{k} \gamma^k \right)$ на множину Θ . У загальному випадку така проєкція може бути знайдена як розв'язок оптимізаційної задачі $\|v_k - \rho_k \gamma^k - v\|^2 \rightarrow \min_{v \in \Theta}$. В одномірному випадку, якщо $\left(v_k - \frac{1}{k} \gamma^k \right) \in \Theta$, то $v_{k+1} = v_k - \frac{1}{k} \gamma^k$.

5. Перевіряємо один з критеріїв зупинки: $|v_{k+1} - v_k| < \varepsilon$, $|L_n(v_{k+1}) - L_n(v_k)| < \varepsilon$ або $|\tilde{\nabla} L_n(v_{k+1})| < \varepsilon$, де $\tilde{\nabla} L_n(v)$ — стохастичний квазіградієнт функції максимальної вірогідності. Якщо він виконується, то вважаємо, що істинне значення параметра $v^* = v_{k+1}$, інакше $k = k + 1$ і повертаємося до кроку 1.

Умови, що гарантують збіжність послідовності $\{v_k\}$ до істинного значення параметра v^* з імовірністю 1, сформульовані у вигляді наступної теореми [3].

Теорема 1. Нехай виконуються такі умови:

$$1) E(\|\gamma^k\|^2 | v_0, \dots, v_k) \leq C, \quad C = \text{const}, \quad k = 0, 1, \dots;$$

$$2) a_k \geq 0, \quad \|b^k\| \leq \beta^k;$$

$$3) \rho_k \geq 0, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 E(\beta^k) < \infty;$$

$$4) \sum_{k=0}^{\infty} \rho_k a_k = \infty \text{ з імовірністю } 1;$$

5) множина Θ опукла і обмежена.

Тоді послідовність $\{v_k\}$ визначена згідно з (3) і (4) сходиться до v^* з імовірністю 1 і $\lim_{k \rightarrow \infty} v_k \in \Theta^*$.

Очевидно, що розглянуті функції і запропоновані в алгоритмі значення параметрів задовольняють умови теореми 1 і таким чином забезпечують збіжність ітераційної процедури до оптимального значення невідомого параметра.

Для валідації результатів оцінювання необхідно порівняти зображення з навчальної вибірки зі створеними на основі моделі новими зображеннями. Для цього можна використати індекс структурної подібності (Structural Similarity Index) або середньоквадратичне відхилення (Mean Square Error).

Отже, отримані результати мають широку прикладну сферу застосування для моделювання стохастичних процесів у галузі економіки, розпізнавання образів, керування телекомунікаційними системами тощо.

СПИСОК ЛІТЕРАТУРИ

1. Кнопов П.С., Самосёнок А.С. О некоторых прикладных задачах марковских случайных процессов с взаимодействием. *Кибернетика и системный анализ*. 2011. № 3. С. 15–32.
2. Самосёнок А. С. Исследование эмпирических оценок параметров гиббсовского распределения, полученных методом максимального правдоподобия. *Кибернетика и системный анализ*. 2013. № 2. С. 178–187.
3. Ермольев Ю.М. Методы стохастического программирования. Москва: Наука, 1976. 240 с.
4. Pflug G.Ch. Optimization of stochastic models. The interface between simulation and optimization. Dordrecht: Kluwer Academic Publishers, 1996. 382 p.

O.S. Samosonok

AN ALGORITHM FOR ESTIMATING THE UNKNOWN PARAMETER OF THE GIBBS DISTRIBUTION BASED ON THE METHOD OF STOCHASTIC QUASI-GRADIENTS

Abstract. The author considers a practical algorithm for estimating an unknown parameter of a mathematical model of a Markov process with local interaction based on the Gibbs distribution. It is proposed to apply the method of stochastic quasi-gradients to the maximum likelihood function, which is constructed from the observations of the implementations of the Gibbs field. The obtained results have a wide application in the modeling of stochastic processes.

Keywords: Gibbs distribution, Maximum Likelihood Estimation, Markov random fields, stochastic quasi-gradient method, parameter estimation.

Надійшла до редакції 27.06.2023