

МЕТОД ПОИСКА ГЛОБАЛЬНОГО ЭКСТРЕМУМА ФУНКЦИЙ ОТ РАЗНОТИПНЫХ ПЕРЕМЕННЫХ

Лбов Г.С.

Институт математики СО РАН, Россия
630090, г. Новосибирск-90, пр. академика Коптюга, 4
lbov@math.nsc.ru

ABSTRACT

We consider a method of search of approximate value of global extremum of function from heterogeneous variables. The method realizes the procedure of adaptive design of trials, using changeable distribution in search space, depending on previous trials results. On each step of designing, to determine the most perspective regions of search, the approximate logical decision function is constructed.

ВВЕДЕНИЕ

Рассматривается функция $y = \varphi(x) = \varphi(x_1, \dots, x_j, \dots, x_n)$, $y \in D_y$, $x \in D$. Обозначим глобальный экстремум (для определенности – глобальный максимум) функции через $y_{max} = \varphi(x^*)$. Пару $\langle x^i, \varphi \rangle$ будем называть испытанием. Под *алгоритмом поиска приближенного значения глобального максимума функции* понимается некоторая процедура последовательного планирования N испытаний с целью нахождения максимально возможного значения функции. Число N обычно невелико, поскольку обычно каждое вычисление функции $\varphi(x)$ в точке x^i предполагает достаточно большие затраты либо на вычисление функции, либо на проведение эксперимента в том случае, когда математическая модель оптимизируемого объекта не задана. Впервые излагаемая ниже идея адаптивного планирования испытаний была предложена и использована при построении алгоритма приближенного поиска глобального экстремума на единичном гиперкубе (случай бинарных переменных). В дальнейшем с использованием этой идеи были предложены соответствующие алгоритмы для случая номинальных переменных (с неупорядоченным набором значений) [2] и для случая как количественных, так и качественных переменных [3]. Заметим, что до сих пор остаются открытыми вопросы о введении достаточно слабых ограничений на класс функций $\varphi(x)$ и об определении оптимальной стратегии планирования N испытаний. Необходимость введения слабых ограничений на класс функций $\varphi(x)$ связана с тем, что при решении прикладных задач априорные сведения о свойствах функции $\varphi(x)$, как правило, отсутствуют. Ниже будет изложен метод оптимизации для случая разнотипных переменных с использованием класса логических решающих функций. Кроме того, рассматриваются вопросы, связанные с введением слабых ограничений на класс функций $\varphi(x)$ в случае разнотипных переменных.

ПОСТАНОВКА ЗАДАЧИ

Пусть задана функция $y = \varphi(x) = \varphi(x_1, \dots, x_j, \dots, x_n)$, $x \in D$, $D = \prod_{j=1}^n D_j$, D_j – множество значений переменной X_j . Для простоты изложения полагаем, что D_j – дискретное множество, представляющее набор упорядоченных или неупорядоченных l_j значений X_j , $|D| = \prod_{j=1}^n l_j = S$.

Кроме того, полагаем, что значения функции на элементах множества D различны между собой. Зададим порядок на этих значениях $(y^1 > \dots > y^i > \dots > y^S)$, полученных в точках

$x^1, \dots, x^i, \dots, x^S$. Под классом эквивалентности понимаем множество функций $\varphi(x)$, для которых в заранее фиксированных точках $x^1, \dots, x^i, \dots, x^S$ соблюдается один и тот же порядок: $y^1 > \dots > y^S$. В дальнейшем любые функции $\varphi_1(x), \varphi_2(x)$ считаем различными только в том случае, если они принадлежат разным классам эквивалентности. Не теряя общности, любой функции $\varphi(x)$ припишем одно и то же множество упорядоченных значений $D_Y = \{y^1, \dots, y^S\}$. Обозначим множество таких функций через F , $\varphi(x) \in F$. Очевидно, что $|F| = S!$.

Введем понятие ε -окрестности максимального значения функции $\varphi(x)$. Заметим, что $y^1 = y_{\max}$, а $y^S = y_{\min}$.

Рассмотрим величину $\varepsilon_l = \frac{y_{\max} - y^l}{y_{\max} - y_{\min}}$, $y^l \in D_Y$, $0 \leq \varepsilon_l \leq 1$. Под ε -окрестностью максимального значения функции $\varphi(x)$ понимается следующее множество:

$$D_{\varepsilon_l} = \{x \mid y^l \leq \varphi(x) \leq y_{\max}\}.$$

Пусть заданы функция $\varphi(x)$ и алгоритм Q планирования N испытаний. Предполагается, что $N \ll S$. Результатом работы алгоритма является набор точек $\{x^{(1)}, \dots, x^{(N)}\}$ и соответствующий набор значений $\{y^{(1)}, \dots, y^{(N)}\}$ такой, что $y^{(1)} > \dots > y^{(N)}$.

При фиксированных функции $\varphi(x)$, алгоритме Q и величине ε_l считаем, что задача решена, если $x^{(1)} \in D_{\varepsilon_l}$. В противном случае задача считается нерешенной. Заметим, что если не

вводить каких-либо ограничений, то любая функция $\varphi(x)$ с вероятностью $1/S!$ может быть предоставлена для решения. Будем говорить, что в этом случае задано равномерное распределение $p_0(\varphi)$ на множестве F . Можно доказать, что при распределении $p_0(\varphi)$ любой алгоритм Q по своей эффективности (по вероятности $P(x^{(1)} \in D_{\varepsilon_l})$) совпадает с методом Монте-Карло, т.е. со случайным выбором N точек в пространстве D с равномерным распределением $p_0(\varphi) = 1/|D|$. Из интуитивных соображений следует, что, по-видимому, в прикладных задачах не все функции равновероятны: существует неравномерное распределение $p(\varphi)$ на множестве функций, т.е. $p(\varphi) \neq p_0(\varphi)$. Вопрос о введении слабых ограничений на класс функций $\varphi(x)$ сводится к разумным соображениям введения неравномерного распределения $p(\varphi)$. Если задано распределение $p(\varphi)$, то можно сформулировать понятие оптимального алгоритма планирования N испытаний: под *оптимальным алгоритмом* понимается алгоритм Q_0 , для которого $P_{Q_0}(x^{(1)} \in D_{\varepsilon_l}) = \max_{Q \in A} P_Q(x^{(1)} \in D_{\varepsilon_l})$, где A – множество алгоритмов планирования N испытаний.

Можно доказать, что оптимальный алгоритм принадлежит классу детерминированных алгоритмов. Введение случайности в выборе точек значительно упрощает алгоритм, однако такой алгоритм отличен от оптимального.

Можно сформулировать и другое понятие оптимального алгоритма. Для любого алгоритма $Q \in A$ потребуем, чтобы $P_Q(x^{(1)} \in D_{\varepsilon_l}) \geq \gamma$, $\gamma = 1 - \delta, \delta \approx 0$. Обозначим через N_Q минимальное число испытаний, необходимое для достижения заданного неравенства. Тогда под оптимальным алгоритмом понимается алгоритм Q_0 , для которого $N_{Q_0} = \min_{Q \in A} N_Q$.

Укажем один из разумных способов введения неравномерного распределения $p(\varphi)$.

ВВЕДЕНИЕ РАСПРЕДЕЛЕНИЯ $p(\varphi)$

Пусть проведено L испытаний ($2 \leq L \leq N$): $\langle x^1, y_1 \rangle, \dots, \langle x^L, y_L \rangle$, где $y_i = \varphi(x^i)$. Рассматривается разбиение n -мерного пространства D на M подмножеств: $\alpha = \{E^1, \dots, E^l, \dots, E^M\}$. Причем

$$E^l = \prod_{j=1}^n E_j^l, E_j \subset D_j \text{ для } j \in I^l; E_j = D_j \text{ для } j \notin I^l, \text{ где } I^l = \{j_1, \dots, j_{m_l}\}, m_l \leq n, \text{ т.е. } \alpha \in \Psi_M.$$

Для каждого E^l определим множество $\mu^l = \{i_1, \dots, i_L\}$ номеров тех из L испытаний, для которых $x^i \in E^l$. Пусть выполняются два ограничения: число $L_l \leq 1$ и нет совпадающих точек из

множества $\{x^{i1}, \dots, x^{iL_i}\}$. Определим для множества E^i максимальное значение функции $y_{max}^i = \max_{i \in \mu_i} y_i$. Для разбиения α упорядочим множества E^1, \dots, E^M в соответствии с порядком $y_{max}^{t_1} > y_{max}^{t_2} > \dots > y_{max}^{t_M}$. Получим упорядоченный набор множеств $\{E^{t_1}, \dots, E^{t_i}, \dots, E^{t_M}\}$. Обозначим через p^i вероятность $P(x^* \in E^{t_i})$, где x^* – точка, в которой достигается глобальный максимум функции $\varphi(x)$.

Для введения ограничения на распределение $p(\varphi)$ используется следующая гипотеза: указанная вероятность p^i зависит от числа проведенных испытаний L , мощности множества $|E^{t_i}|$, занимаемого номера i множества E^{t_i} в указанном выше порядке. При фиксированном $|E^{t_i}|$ чем меньше номер i , тем выше вероятность p^i ; при фиксированном номере i чем больше $|E^{t_i}|$, тем больше p^i . Функцию $p^i = \chi(|E^{t_i}|, i, L)$ определим ниже. В дальнейшем будем полагать, что $L^i \geq L^*$, где L^* – некоторый параметр.

Введем распределение $p(x) = P(x=x^*)$, $x \in D$, следующим образом: если $x \in E^{t_i}$, то $p(x) = p^i / |E^{t_i}|$. Очевидно, что при $p(\varphi) = 1/S!$ вероятность p^i не зависит ни от номера i , ни от числа проведенных испытаний L , а только от $|E^{t_i}|$, т.е. $p^i = |E^{t_i}| / |D|$. Тогда $p(x) = p_0(x) = 1/|D|$, $x \in D$. Другими словами, при $p(\varphi) = 1/S!$ проведенные испытания не дают информации о местоположении глобального экстремума. Вычислим энтропию $H_0 = - \sum_{x \in D} p_0(x) \log p_0(x) =$

$\log S$, определяющую максимальную меру неопределенности местоположения точки x^* . Задавая неравномерное распределение $p(x)$, тем самым вводим неравномерное распределение $p(\varphi)$. Отметим, что чем больше отклонение $p(x)$ от равномерного, тем меньше энтропия H . При фиксированном числе проведенных испытаний L , $L \leq N$, чем меньше величина $\rho(L) = H(L)/H_0$, где $H(L) = - \sum_{x \in D} p(x) \log p(x)$, тем больше отклонение распределения $p(\varphi)$ от равномерного (тем “сильнее” гипотеза). Ясно, что при увеличении числа испытаний L величина $\rho(L)$ должна уменьшаться.

Исходя из указанных выше общих соображений, приведем пример реализации данной идеи.

АЛГОРИТМ ADAPT

Алгоритм ADAPT реализует последовательную процедуру планирования N испытаний и в общем случае предназначен для поиска приближенного значения глобального экстремума функции от разнотипных переменных.

Общее число испытаний N разбивается на R групп $N = N_1 + \dots + N_i + \dots + N_R$.

Обозначим через $N^v = \sum_{i=1}^v N_i$ число испытаний, проведенных за v шагов поиска (под шагом поиска понимается проведение N_i испытаний), $v=1, \dots, R$. До проведения испытаний распределение $p(x) = p_0(x) = 1/|D|$. Первая группа испытаний планируется таким образом, чтобы для любого $E \subset D$ число испытаний, соответствующее множеству E , примерно равнялось $N_1 |E| / |D|$. Другими словами, точки x^1, \dots, x^{N_1} планируем таким образом, чтобы они были максимально равномерно распределены по множеству D . Пусть проведено N^v испытаний ($v=1, \dots, R-1$) и получена соответствующая таблица $v^v = \{x^i, y^i\}$, $i=1, \dots, N^v$. С помощью алгоритма LRP [4] строим наилучшую регрессионную функцию \bar{f}^v в классе логических

решающих функций. Функции \tilde{f}^v соответствует разбиение $\alpha_v = \{E_v^1, \dots, E_v^t, \dots, E_v^{M_v}\}$. Если функция $\rho(N^v)$ задана (способ задания функции $\rho(N^v)$ будет определен ниже), можно определить вероятности $p_t^v, t=1, \dots, M_v$.

Для этого используется вспомогательная функция $\lambda_b(z) = (b+1)/(1+bz)^2, 0 \leq z \leq 1, 0 \leq b \leq \infty$. Выбор этой функции определяется следующими ее свойствами:

- 1) при $b=0$ получаем $\lambda_b(z)=1$;
- 2) при $b > 0$ получаем монотонно убывающую функцию по z ;
- 3) чем больше значение параметра b , тем больше скорость убывания функции по z ;
- 4) $\int_0^1 \lambda_b(z) dz = 1$ при любом b .

Данные свойства функции $\lambda_b(z)$ позволяют формализовать вышеуказанную гипотезу. Отметим, что в качестве функции $\lambda_b(z)$ можно выбрать любую функцию, удовлетворяющую этим свойствам. Зададим функцию $\lambda^u = \int_0^u \lambda_b(z) dz = (b+1)u/(1+bu), 0 \leq u \leq 1$.

Определим вероятность p_t^v , соответствующую множеству E_t^v . По оси z откладываем относительные мощности $E_t^v/|D|$. Причем первоначально откладываются относительная мощность наилучшего множества E_t^{v1} , которому соответствует y_{max}^{t1} , затем относительная мощность второго по порядку множества E_t^{v2} , которому соответствует y_{max}^{t2} , и т.д. Таким образом, отрезок $[0,1]$ на оси z разбивается на M_v отрезков. Вероятность p_t^v определяется по формуле

$$p_t^v = \lambda^{u_t+1} + \lambda^{u_t} = \int_0^{u_{t+1}} \lambda_b(z) dz - \int_0^{u_t} \lambda_b(z) dz = \frac{(b+1)(u_{t+1} - u_t)}{(1+bu_{t+1})(1+bu_t)}, t=1, \dots, M_v$$

Из изложенного ясно, что чем больше мощность множества E_t^v и меньше номер i , определяющий номер этого множества в порядке $\{E_v^1, \dots, E_v^t, \dots, E_v^{M_v}\}$, тем больше вероятность p_t^v (в соответствии с указанной выше гипотезой). Зададим линейную зависимость параметра b от числа проведенных испытаний N^v , т.е. $b^v = N^v b_{max}/N$. Величина b_{max} определяется из следующих соображений. После проведения всех испытаний ($N^R = N$) в соответствии с гипотезой можно указать область $E(\gamma)$ такую, что вероятность нахождения точки x^* равна $p(\gamma) \approx 1$. При этом область $E(\gamma)$ достаточно мала, т.е. $\gamma = E(\gamma)/|D| \approx 0$, например, $p(\gamma) = 0,95, \gamma = 0,05$. Величина b_{max} определяется из соотношения $p(\gamma) = \int_0^\gamma \lambda_u(z) dz = \frac{(b_{max}+1)\gamma}{1+b_{max}\gamma}$. Очевидно, что величина $b_{max} = (p(\gamma) - \gamma)/(\gamma - \gamma p(\gamma))$ с учетом линейной зависимости $b^v = N^v b_{max}/N$ определяет функцию $\rho(L) = H(L)/H_0$, где $L = N^v$.

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ АЛГОРИТМА ОПТИМИЗАЦИИ

Предложенный алгоритм применялся для функции от разнотипных переменных. Исследование проводилось с помощью процедуры статистического моделирования с использованием модификации функции Шекеля:

$$f(x_1, x_2) = -x_2 \sum_{i=1}^{10} \frac{1}{(k_{i1}(x_1 - a_{i1})^2 + c_{i1})} - (1 - x_2) \sum_{i=1}^{10} \frac{1}{(k_{i2}(x_1 - a_{i2})^2 + c_{i2})}$$

где $0 \leq x_1 \leq 10, x_2 \in \{0,1\}$, а коэффициенты – случайные числа с равномерным распределением на интервалах $0 \leq a_{i1}, a_{i2} \leq 10, 1 \leq k_{i1}, k_{i2} \leq 3, 0.1 \leq c_{i1}, c_{i2} \leq 0.3$. Таким образом, функция

$f(x_1, x_2)$ является случайной функцией от разнотипных переменных (X_1 – количественная, X_2 – бинарная переменная).

Для каждой из 100 реализаций данной функции с помощью предложенного алгоритма найдено приближенное значение глобального минимума f_{min}^* путем адаптивного планирования 200 вычислений функций (испытаний). Затем определена погрешность в вычислении глобального минимума величиной

$$\delta = \frac{f_{min}^* - \tilde{f}_{min}}{\tilde{f}_{max} - \tilde{f}_{min}} 100\%,$$

где \tilde{f}_{min} и \tilde{f}_{max} – приближенные значения минимума и максимума реализации функции на основе ряда вычислений по сетке двумерной области с достаточно малым шагом по переменной X_1 . Средняя погрешность $\bar{\delta}$ за 100 реализаций функций равна 0,78%.

Работа проделана при поддержке гранта РФФИ №98-01-00673.

ЛИТЕРАТУРА

1. Лбов Г.С. Выбор эффективной системы зависимых признаков. -Вычислительные системы, 1965, вып.19, с. 21-34
2. Lbov G.S. Training for extremum determination of function of variables measured in names scale. In: Second Intern. Conf. on artificial intelligence. London, 1972,p.418-423
3. Lbov G.S. Algorithm of searching the global extremum of function of variables measured in different scales. - In: Proc. 9-th Intern. Conf. on optimization techneques, IFIP. 1979, pt.2 New York a.o. 1980, p. 87-95
4. Лбов Г.С., Старцева Н.Г. Логические решающие функции и вопросы статистической устойчивости решений. Изд-во Института математики СО РАН, Новосибирск, 1999, 211 с.