

УДК 621.391:519.2.

АДАПТИВНЫЙ ПОИСК ЭКСТРЕМУМА ФУНКЦИИ ОТ ПЕРЕМЕННЫХ, ЗАМЕРЕННЫХ В ШКАЛЕ НАИМЕНОВАНИЙ

Г.С. Лбов

§ 1. Постановка задачи

Рассматривается следующая задача. Пусть имеется набор из n переменных $X_1, X_2, \dots, X_i, \dots, X_n$. Каждая из переменных принимает конечное множество значений $X_i = (x_{i1}, \dots, x_{i2}, \dots, x_{i\ell_i})$. Переменные являются величинами, замеренными в шкале наименований [1]. Другими словами, каждое такое множество значений представляет собой список имен некоторых объектов. Это означает, что пространство, построенное на этих переменных, не является метрическим.

Образуется набор из n элементов $x = (x_1, x_2, \dots, x_i, \dots, x_n)$, по одному элементу из каждого множества. Число различных наборов равно

$$N = \prod_{i=1}^n \ell_i.$$

Считается, что для каждого такого набора элементов экспериментальным путем можно определить значение критерия качества этого набора $F(x)$. Требуется найти такой набор $x^* = (x_1^*, \dots, x_i^*, \dots, x_n^*)$, при котором функция $F(x^*)$ принимала бы экстремальное значение. В такой постановке задача, вообще говоря, может быть решена только путем полного перебора. Но, как правило, при решении практических задач такой перебор невозможен из-за большого числа вариантов N и больших затрат на проведение эксперимента. Можно перебрать лишь ограниченное число T различных

вариантов ($T \ll N$). В этом случае, если не делать никаких предположений об ограничениях на класс функций, любой из алгоритмов поиска экстремума функции $F(x)$ одинаково предпочтителен.

В данной работе предлагается алгоритм, основанный на идее адаптивного случайного поиска. Кратко эта идея заключается в следующем: организация последующих испытаний зависит от результатов всех предыдущих. Происходит своего рода обучение на экстремум. Для этого общее число испытаний разбивается на ряд групп

$$T = r^{(1)} + r^{(2)} + \dots + r^{(v)} + \dots + r^{(k)}$$

и при испытаниях $(y+1)$ группы используется такое предположение: для любой пары значений $\{x_{i,y}, x_{i,y}^*\}$ переменной X_i вероятность $\rho^{(y)}\{x_{i,y} = x_{i,y}^*\} > \rho^{(y)}\{x_{i,y} = x_{i,y}^*\}$, если хотя бы один набор x , содержащий $x_{i,y}$, лучше (по критерию $F(x)$) любого набора x , содержащего $x_{i,y}^*$. Разность между указанными вероятностями тем больше, чем больше относительное число проведенных испытаний к данному моменту поиска

$$\bar{r}_y = \frac{1}{T_y} \sum_{j=1}^y r^{(j)}$$

Испытания $(y+1)$ группы проводятся так, чтобы значения $x_{i,y}$ переменной X_i включались в наборы пропорционально вероятности $\rho^{(y)}\{x_{i,y} = x_{i,y}^*\}$, которая в свою очередь изменяется по мере проведения поиска. В начале поиска при отсутствии априорной информации о предпочтении одного значения переменной X_i другому вероятности выбора этих значений полагаются равными $\frac{1}{\ell_i}$.

Чем больше число испытаний T , тем более "осторожным" становится поиск экстремума и тем шире класс решаемых задач.

По своей постановке рассматриваемая задача не сводится к известным задачам дискретного программирования [2].

В качестве примера задач, для решения которых предполагается использовать описываемый ниже алгоритм, может служить задача выбора наиболее эффективной подсистемы зависимых признаков при распознавании образов. Действительно, множество всех возможных вариантов в этом случае представляет собой множество всех решений о включении того или иного признака в эффективную

подсистему. Переменная X_i принимает два значения: либо i -й признак включен в подсистему признаков, либо не включен. Число вариантов в этом случае равно $N = 2^n$.

В качестве критерия F могут использоваться, например, суммарные потери, связанные как с измерением признаков, так и с ошибками распознавания.

Другим примером задач может быть задача оптимизации технологических процессов, если при этом встречаются переменные, измеренные в шкале наименований.

Идея адаптивного случайного поиска была использована ранее в алгоритме СНА [3] для выбора m признаков из n ($N = C_n^m$). Этот алгоритм оказался достаточно эффективным при решении ряда практических задач распознавания образов из области медицины, геологии, социологии, экономики, психологии.

§ 2. Описание алгоритма

Введем следующие понятия:

1. Пространство позиций. Обозначим его через Z . Пространство позиций строится на переменных $x_1, x_2, \dots, x_i, \dots, x_n$, поставленных в соответствие переменным $X_1, X_2, \dots, X_i, \dots, X_n$. Переменная X_i принимает одно из целочисленных значений $\{1, 2, \dots, \beta_i, \dots, m_i\}$, которые назовем позициями. Число позиций m_i равно числу значений переменной X_i в данный момент поиска ($m_i \leq \ell_i$)*.

После проведения каждой группы экспериментов между позициями переменной X_i и значениями переменной X_i устанавливается взаимоднозначное соответствие. Для этого фиксируется наилучший набор $x^{(1)}$, содержащий значение $x_{i,1}$, наилучший набор $x^{(2)}$, содержащий значение $x_{i,2}$, и т.д. Упорядочивая наборы $x^{(1)}, x^{(2)}, \dots, x^{(m_i)}$ по критерию $F(x)$, мы тем самым упорядочиваем значения переменной X_i . Первой позиции ставится в соответствие то значение переменной, для которого получен наилучший результат по критерию $F(x)$, второй позиции — значение, для которого получен следующий по величине результат, и т.д.

*) По мере проведения поиска число используемых значений переменной X_i уменьшается путем исключения менее перспективных значений переменной. Выбор числа m_i будет рассмотрен ниже.

2. Набор переводных векторов. Обозначим его через $\alpha = (\alpha_1, \dots, \alpha_{i_1}, \dots, \alpha_n)$.

Вектор $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{i\beta}, \dots, \alpha_{i\ell_i})$ устанавливает соответствие между значениями переменной X_i и позициями переменной X_i . Компонента этого вектора $\alpha_{i\beta}$ представляет собой то значение переменной X_i , которое занимает β -я позиция в данный момент поиска.

В начале поиска, если нет априорной информации о предпочтении одного значения переменной X_i другому, вектор α_i имеет вид

$$\alpha_i = (x_{i1}, x_{i2}, \dots, x_{i\beta}, \dots, x_{i\ell_i}).$$

Так как набор векторов обновляется после каждой группы экспериментов, обозначим его через $\alpha^{(y)}$.

3. Матрица рекомендаций $B^{(y)}$. Эта матрица представляет собой набор рекомендаций в пространстве позиций для проведения экспериментов y -й группы.

$$B^{(y)} = \{x_i^{(y)}\} \text{ для } i=1, \dots, n; y=1, \dots, r^{(y)}$$

4. Матрица планирования $\phi^{(y)}$. Она получается из матрицы рекомендаций $B^{(y)}$ с помощью набора векторов $\alpha^{(y)}$.

Матрица планирования имеет вид:

$$\phi^{(y)} = \{x_i^{(y)}\} \text{ для } i=1, \dots, n; y=1, \dots, r^{(y)}$$

5. Функция уверенности P . Эта функция вводится как количественная мера предпочтения одного значения переменной другому. Эта функция имеет вид:

$$P = \{P_\beta(b_i, m_i)\} \text{ для } i=1, \dots, n; \beta=1, \dots, \ell_i, \text{ где}$$

$P_\beta(b_i, m_i)$ — вероятность того, что значение переменной, которое в данный момент поиска занимает β -я позиция, есть значение X_i^* . Выбор параметра b_i будет рассмотрен ниже.

Функция уверенности P задается в пространстве позиций.

Согласно предположению, описанному в § 1, функция $P_\beta(b_i, m_i)$ должна обладать следующими свойствами:

1) в начале поиска при отсутствии априорных сведений о предпочтении одного значения переменной другому

$$P_\beta(b_i, m_i) = \frac{1}{m_i};$$

2) функция $P_\beta(b_i, m_i)$ должна монотонно убывать с увели-

чением номера позиции;

3) при увеличении относительного числа проведенных экспериментов функция уверенности должна возрастать на первых позициях за счет позиций с большими номерами;

$$4) \sum_{\beta=1}^{m_i} P_\beta(b_i, m_i) = 1.$$

Была использована функция

$$f(x) = \frac{b+1}{(1-bx)^2},$$

которая приводится в работе [4]. Ее поведение в зависимости от изменения параметра b проиллюстрировано на рис. 1.

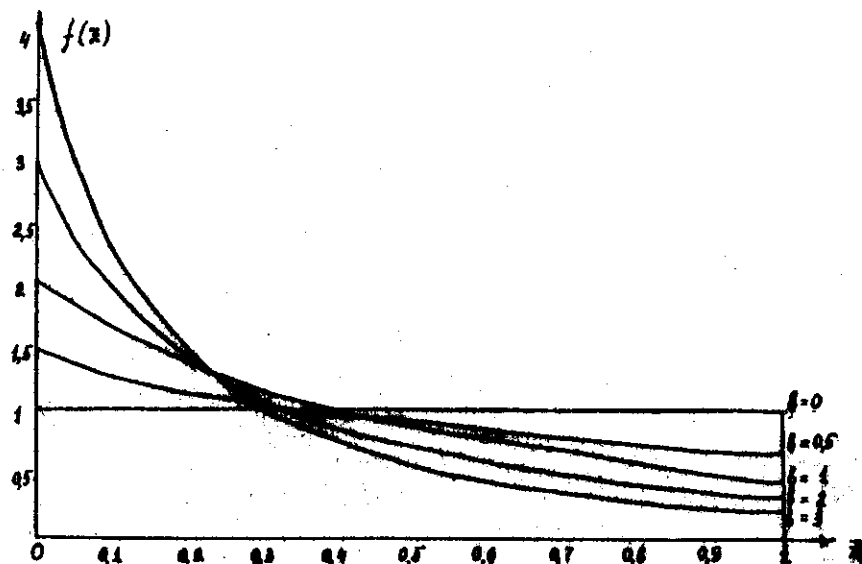


Рис. 1. Функция $f(x)$

Для любого значения параметра b из интервала $(0, \infty)$

$$\int_0^1 f(x) dx = 1.$$

Функция уверенности задается следующим образом

$$P_\beta(b_i, m_i) = \int_{x=1}^{m_i} \frac{b_i+1}{(1+b_i x)^2} dx = \frac{b_i+1}{b_i + \frac{m_i}{\beta}} - \frac{b_i+1}{b_i + \frac{m_i}{\beta-1}}. \quad (1)$$

Функция (1) обладает сформулированными выше свойствами. Действительно, если $b_i = 0$, то все m_i позиций переменной x_i имеют одинаковое значение функции уверенности, равное $\frac{1}{m_i}$. При

$b_i > 0$ функция $\rho_i(b_i, m_i)$ монотонно убывает с увеличением номера позиции. Если увеличивать параметр b_i по мере накопления числа экспериментов, то значения переменной x_i , занимающие первые позиции, будут предпочтительнее. Для любого значения b_i :

$$\sum_{j=1}^{m_i} \rho_j(b_i, m_i) = 1.$$

Так как функция уверенности зависит от числа групп проведенных экспериментов, обозначим её через $\rho_i^{(v)}(b_i, m_i)$.

Если теперь число экспериментов, соответствующих тому или иному значению переменной x_i , делать пропорциональным значению функции уверенности, то по мере накопления числа экспериментов при дальнейшем поиске будут чаще включаться всё более перспективные значения переменных $x_1, \dots, x_n$. Число экспериментов $r_{i,p}^{(v)}$ для значения переменной, занимающего p -ю позицию, делается равным ближайшему целому от $r^{(v)} \cdot \rho_p^{(v)}(b_i, m_i)$, то есть

$$r_{i,p}^{(v)} = \text{цел}\{r^{(v)} \cdot \rho_p^{(v)}(b_i, m_i)\}. \quad (2)$$

На каждом этапе поиска устанавливается взаимоднозначное соответствие между точками пространства Z , построенного на переменных $x_1, \dots, x_n$, и пространства X , построенного на переменных $\lambda_1, \dots, \lambda_n$. Это соответствие устанавливается с помощью набора векторов $\alpha^{(v)}$. Поэтому если в пространстве Z выбирать множество из $r^{(v)}$ точек в соответствии с законом распределения, задаваемым функцией уверенности, удовлетворяя при этом условию (2), то и в пространстве

X этому множеству будет соответствовать множество точек, которое также удовлетворит условию (2). Это дает нам возможность получить множества точек в пространстве Z (матрицы $B^{(1)}, B^{(2)}, B^{(3)}, \dots, B^{(k)}$) до проведения экспериментов. Для этого необходимо задать ЭВМ лишь следующие данные: число переменных n , число значений для каждой из переменных $\ell_1, \ell_2, \dots, \ell_i, \dots, \ell_n$ и число экспериментов T .

Опишем процедуру получения матриц $B^{(1)}, B^{(2)}, \dots, B^{(v)}, \dots, B^{(k)}$. Прежде всего необходимо определить функции уверенности $\rho^{(v)}(b_i, m_i)$, используя формулу (1). Задание коэффициента b_i и числа позиций m_i происходит следующим способом.

Для переменной x_i можно вычислить энтропию, используя известное соотношение

$$H^{(v)}(b_i, m_i) = - \sum_{p=1}^{m_i} \rho_p^{(v)}(b_i, m_i) \cdot \log \rho_p^{(v)}(b_i, m_i). \quad (3)$$

Далее вычисляется энтропия для всего пространства позиций. Эта энтропия равна сумме энтропий по каждой переменной, поскольку выбор позиций по каждой из координат производится независимо, то есть

$$H^{(v)} = \sum_{i=1}^n H^{(v)}(b_i, m_i).$$

Величина $H^{(v)}$ является количественной мерой нашего незнания о том, какой из N возможных вариантов соответствует наилучшему значению критерия $F(\alpha)$. В начале поиска, если нет априорной информации о предпочтении одного значения переменной другому, эта энтропия равна

$$H^{(0)} = \sum_{i=1}^n H_i^{(0)} = \sum_{i=1}^n \log \ell_i = \log N.$$

Обозначим через $\tau_v = \frac{t_v}{T}$ относительное число проведенных испытаний, а через $T_v = T - t_v$ — число оставшихся испытаний. Потребуем, чтобы энтропия $H^{(v)}$ монотонно убывала до нуля по мере возрастания τ_v до единицы. Для этого вводится функция (рис. 2):

$$H^{(v)} = \log \lambda_v + H^{(0)}, \quad \text{где}$$

$$\lambda_v = \frac{\kappa \tau_v^2}{2} + [\alpha^{-H^{(0)}} - (\frac{\kappa}{2} + 1)] \tau_v + 1. \quad (4)$$

Величина α есть основание логарифма. Параметр κ определяет степень адаптации. Чем меньше κ , тем меньше степень адаптации. Значение этого параметра должно выбираться из интервала $(-2, 2)$ для того, чтобы $0 \leq \lambda_v \leq 1$.

При изменении величины τ_v от нуля до единицы величина λ_v изменится от единицы до $\alpha^{-H^{(0)}}$, а энтропия $H^{(v)}$

монотонно уменьшается от $H^{(0)}$ до нуля.

Ниже будет приведено условие, от которого зависит выбор конкретного значения параметра k .

После проведения t , испытаний энтропия уменьшается по сравнению с первоначальной в $\frac{H^{(t)}}{H^{(0)}}$ раз. Будем считать, что энтропия по каждой из переменных также уменьшается во столько же раз, т.е.

$$\frac{H_i^{(t)}}{H_i^{(0)}} = \frac{H^{(t)}}{H^{(0)}}$$

Тогда

$$H_i^{(t)} = H^{(t)} \cdot \frac{H_i^{(0)}}{H^{(0)}} = (\log \lambda_t + H^{(0)}) \cdot \frac{\log \ell_i}{\log N} \quad (5)$$

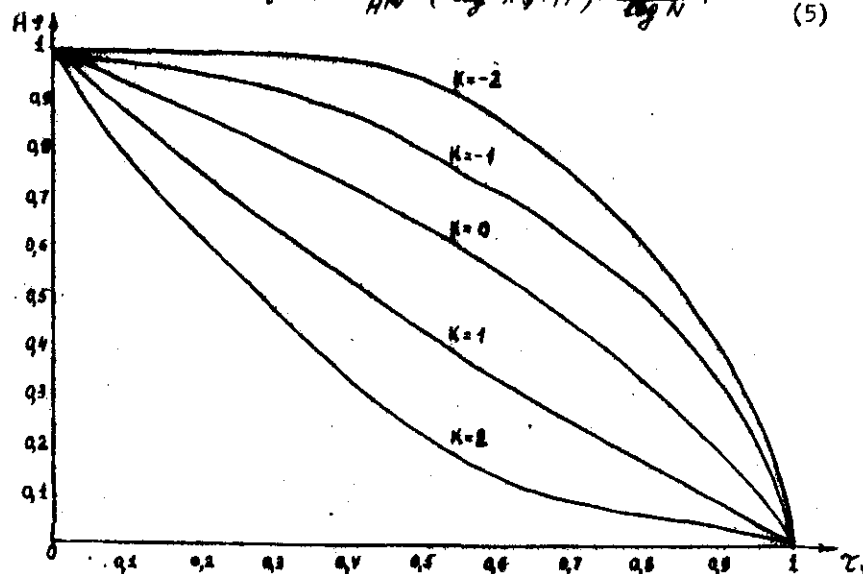


Рис. 2. Зависимость энтропии от относительного числа проведенных экспериментов (при $H^{(0)} = 1$).

Приравняем энтропию $H^{(t)}(\ell_i, m_i)$ и энтропию $H_i^{(t)}$, определяемые соотношениями (3) и (5):

$$H^{(t)}(\ell_i, m_i) = H_i^{(t)} \quad (6)$$

Определим число позиций m_i . Для этого определяется максимальный номер позиции, для которой из оставшегося числа ис-

пытаний T_i можно выделить хотя бы одно. Этот номер определяется следующим образом. Сначала полагается $m_i = \ell_i$ и из уравнения

$$P_{\ell_i}(\ell_i, \ell_i) \cdot T_i = 1$$

определяется значение коэффициента ℓ_i . Обозначим его через $\ell_i \text{ max}$. Если при этом

$$H^{(t)}(\ell_i \text{ max}, m_i) > H_i^{(t)}, \text{ то}$$

m_i полагается равным $\ell_i - 1$. Для нового значения m_i проводится аналогичная проверка. Эта процедура повторяется до тех пор, пока не выполнится неравенство

$$H^{(t)}(\ell_i \text{ max}, m_i) \leq H_i^{(t)} \quad (7)$$

За число m_i принимается первое число m , при котором начинает выполняться неравенство (?).

При найденном числе m_i коэффициент ℓ_i находится из уравнения (6) методом дихотомии.

Определив ℓ_i и m_i , можно вычислить вероятности $P_{\ell_i}(\ell_i, m_i)$ для $i=1, \dots, m_i$, используя выражение (1).

Далее находится число $n^{(v+1)}$ испытаний, включаемых в $(v+1)$ группу. Это число должно быть такое, чтобы, получив результаты экспериментов $F_1, \dots, F_{n^{(v+1)}}$, можно было упорядочить значения каждой переменной X_i по величине F . Для того, чтобы поиск экстремума был максимально "осторожным", $n^{(v+1)}$ должно быть возможно меньшим числом. Для этого потребуем, чтобы то значение переменной, которое в данный момент занимает последнюю позицию, включалось бы в испытание один раз. Из этих замечаний следует, что

$$n^{(v+1)} = \max_{i=1, \dots, n} \{ \text{окр } n_i^{(v+1)} \}, \text{ где}$$

$$\text{окр } \left\{ \frac{n_i^{(v+1)}}{P_{m_i}(\ell_i, m_i)} \right\} \text{ представляет собой ближайшее целое от}$$

Число испытаний $n_i^{(v+1)}$ для $i=1, \dots, m_i$ и $i=1, \dots, n$ определяется с помощью выражения (2).

Далее, в соответствии с законом распределения, задаваемым функцией уверенности, выбирается множество из $n^{(v+1)}$ точек

в n -мерном пространстве позиций, причем выбор этих точек осуществляется таким образом, чтобы число точек, соответствующих β -й позиции переменной Z_i , равнялось $n_i^{(\beta)}$. Координаты этих точек представляются в виде матрицы $S^{(\beta)}$ порядка $(n_i^{(\beta)} \times n)$.

Процедура получения матрицы $S^{(\beta)}$ повторяется несколько раз. Из полученных матриц выбирается та матрица, которой соответствует наименьшее число совпадающих строк (совпадающих точек). Если относительное число совпадающих строк превышает некоторый порог, то уменьшается степень адаптации (уменьшается значение параметра κ , начиная с $\kappa = 2$), и процесс получения матриц $S^{(1)}, S^{(2)}, \dots, S^{(\kappa)}$ проводится заново.

Если при уменьшении κ вплоть до $\kappa = -2$ не удалось получить все матрицы $S^{(1)}, \dots, S^{(\kappa)}$, то происходит останов машины и выдача полученных матриц и число нераспределенных испытаний.

Дальнейшая работа (планирование эксперимента) сводится к переводу матрицы $S^{(\beta)}$ в матрицу $S^{(\alpha)}$ с помощью набора векторов $\alpha^{(i)}$ и обновлении $\alpha^{(i)}$ на основе полученных результатов $F_1, \dots, F_n$. Получение матрицы $S^{(\alpha)}$ может производиться и без применения ЭВМ. Это обстоятельство особенно важно, поскольку экспериментатор пока не всегда может использовать ЭВМ.

Л и т е р а т у р а

1. Суппес П., Зинес Дж. Основы теории измерений. Сб. "Психологические измерения", Москва, Изд. "Мир", 1967.
2. Корбут А.А., Финкельштейн Ю.Ю. Дискретное программирование. Москва, Изд-во "Наука", 1969.
3. Лбов Г.С. Выбор эффективной системы зависимых признаков. "Вычислительные системы", Новосибирск, вып. 19, 1965.
4. Бусленко Н.П., Шрейдер Ю.А. Метод статистических испытаний. Москва, ФМ, 1961.

Поступила в редакцию
27.I.1971.