УДК 62-50

МЕТОДЫ КУСОЧНО-ЛИНЕЙНОЙ АППРОКСИМАЦИИ И ИХ ИСПОЛЬЗОВАНИЕ В ЗАДАЧАХ УПРАВЛЕНИЯ[1]

© Бауман Е.В., Гольдовская М.Д., Дорофеюк Ю.А.

Институт проблем управления РАН

Abstract. This paper is dedicated to the methods of solution piecewise-linear approximation problem, which were developed on the base of general approach to range data analysis. The main idea of piecewise approximation of complex dependence is to divide the in-parameter space into areas, so that within the bounds of each of these areas it will be able to approximate the complex dependence in the whole space with linear function.

1. Постановка задачи кусочно-линейной аппроксимации

1.1. **Случай чёткой классификации.** Рассматривается задача построения модели зависимости выходного показателя y (объясняемой переменной) от вектора входных показателей $x = \left(x^{(1)}, ..., x^{(k)}\right) \in X = \mathbf{R}^k$ (объясняющих переменных). Модель строится по выборке из n объектов, каждый из которых описывается вектором $(y_t, x_t) = \left(y_t, x_t^{(1)}, \ldots, x_t^{(k)}\right) \in \tilde{X} = \mathbf{R}^{k+1}$.

Классическая схема кусочно-линейной аппроксимации состоит в следующем [1]:

Пространство с помощью одного из алгоритмов автоматической классификации [2] разбивается на r классов $(H_1, ..., H_r)$. Затем в каждом классе по методу наименьших квадратов строится линейная регрессия выходного показателя y от вектора входных показателей x. В i-м классе находятся такой вектор коэффициентов $\mathrm{c}_i = \left(c_i^{(1)}, ..., c_i^{(k)}\right)$ и константа d_i линейной функции$((c_i, x) + d_i) = d_i + \sum\limits_{j=1}^{k} c_i^j x^j$, которые минимизируют функционал $K_i = \sum\limits_{x_t \in H_i} (y_t - ((c_i, x_t) + d_i))^2$. Задача кусочно-линейной аппроксимации состоит в нахождении такого разбиения на классы, для которого сумма квадратов невязок по моделям всех классов была бы минимальна. Другими словами, необходимо найти такую классификацию $(H_1, ..., H_r)$ и такие векторы коэффициентов c_i и константы d_i, для которых функционал $I = \sum\limits_{i=1}^{r} \sum\limits_{x_t \in H_i} (y_t - ((c_i, x_t) + d_i))^2$ принимал бы минимальное значение или функционал $J_{\text{КА}} = -I = -\sum\limits_{i=1}^{r} \sum\limits_{x_t \in H_i} (y_t - ((c_i, x_t) + d_i))^2$принимал бы максимальное значение. Последний функционал является частным случаем функционала классификационного анализа общего вида [2]: $J(H, A) = \sum\limits_{x \in X} \sum\limits_{i=1}^{r} K(x, \alpha_i)\, \varphi(h_i(x))$.

[1]Работа выполнена при частичной финансовой поддержке РФФИ, проекты 08-07-00349-а, 08-07-00347-а.

Классификацию $H = (H_1, ..., H_r)$ будем задавать через вектор-функцию принадлежностей $H(x) = (h_1(x), ..., h_r(x))$. Тогда функционал I можно переписать в следующих двух эквивалентных записях:

$$I_1 = \sum_{i=1}^{r} \sum_{t=1}^{n} \left[y_t - ((c_i, x_t) + d_i)\right]^2 h_i(x), \qquad (1)$$

$$I_2 = \sum_{t=1}^{n} \left[y_t - \sum_{i=1}^{r} ((c_i, x_t) + d_i)\ h_i(x)\right]^2. \qquad (2)$$

Когда классификация чёткая (т.е. каждый объект однозначно относится к одному из классов), функционалы (1) и (2) совпадают. Однако интерпретируются они поразному. При минимизации функционала (1) линейные модели зависимостей строятся в каждом классе в отдельности, а затем суммируются квадраты отклонения ошибок. В функционале (2) выражение $\sum\limits_{i=1}^{r} ((c_i, x_t) + d_i)\ h_i(x)$ можно считать кусочно-линейной моделью выходного показателя y.

Решению задачи построения кусочно-линейной зависимости в такой постановке посвящены, например, публикации [1, 3, 4].

Основной сложностью решения данной задачи является то, что при минимизации функционалов (1) и (2) по классификации и по коэффициентам модели решающие правила оптимальной классификации $H = (H_1, ..., H_r)$ записываются в терминах не только входных, но и выходного показателя. Это существенно уменьшает возможность использования построенной модели для прогноза. Приходится использовать различные варианты проекции классификации в расширенном пространстве $\tilde{X}$ на пространство X входных показателей. Подобное проектирование областей на пространство меньшей размерности приводит к появлению в пространстве входных показателей зон, в которых одновременно могут действовать модели как одного, так и другого класса. Именно исходя из этого, при построении кусочных моделей по существу возникает размытость между классами.

1.2. **Случай размытой классификации.** Рассмотрим возможность использования в задаче кусочной аппроксимации размытой классификации. Будем задавать классификацию через вектор-функцию принадлежностей $H(x) = (h_1(x), ..., h_r(x))$, удовлетворяющую ограничению:

$$\sum_{l=1}^{r} h_i(x) = 1; \quad h_i(x) \geq 0,\ \ x \in X,\ \ l = 1, ..., r\,. \qquad (3)$$

В данном случае функционалы (1) и (2) не совпадают, и их минимизация приводит к разным результатам. Так как функционал $I_1 = I_1(c_i, d_i, i = 1, ..., r; H(x))$ линеен по $H(x)$, то функционал $\hat{I}_1 = -\min\limits_{c_i, d_i} I_1(c_i, d_i, i = 1, ..., r; H(x))$ является выпуклым по $H(x)$. Отсюда следует, что оптимальная классификация $H(x)$ лежит на границе допустимой области, т.е. в ограничении все $h_i(x)$ равны либо 1, либо 0, что соответствует четкой классификации. Следовательно, в оптимальной кусочной аппроксимации для функционала (1) при ограничениях (3) классификация - чёткая.

Рассмотрим функционал (2). Достаточно легко показать, что, если на $h_i(x)$ не накладывать дополнительные ограничения, то за счёт большого количества степеней свободы можно всегда точно аппроксимировать все значения y на данной выборке.

Для того чтобы использовать функционал (2) в размытом варианте, необходимо ограничить класс разрешённых функций $h_i(x)$. В соответствии с общей методикой обобщённого среднего [2] для получения кусочной аппроксимации с размытой классификацией функционал (1) необходимо модифицировать следующим образом:

$$I_3 = \sum_{i=1}^{r} \sum_{t=1}^{n} \left[y_t - ((c_i, x_t) + d_i)\right]^2 \varphi\left(h_i(x)\right). \tag{4}$$

Здесь $\varphi(h)$ – одна из ниже перечисленных функций: 1) $\varphi_1(h) = h$ приводит к чёткой кусочной аппроксимации (функционалы (1) и (2) совпадают); 2) $\varphi_2(h) = (h)^t, t > 1$, приводит к размытой кусочной аппроксимации с функцией принадлежности $h_i(x)$; 3) $\varphi_3(h) = t - \sqrt{t^2 - (2t-1)h}$, $t > 1$ приводит к кусочной аппроксимации с классификацией с размытыми границами (размываются лишь границы классов).

Поскольку функционал (3) является частным случаем функционала классификационного анализа общего вида, то для его оптимизации можно использовать общий алгоритм классификационного анализа [2] - последовательное применение двухэтапной процедуры**:** 1) фиксируется вектор-функция $H(x)$, для неё находятся оптимальные значения коэффициентов модели $c_i, d_i, i = 1, ..., r$; 2) фиксируются коэффициенты $c_i, d_i, i = 1, ..., r$ и для них находится оптимальная вектор-функция $H(x)$. Сходимость этого алгоритма следует из сходимости алгоритма [2].

Как и для чёткого случая, недостатком такого подхода является то, что в решающие правила аппроксимации входят не только входные показатели, но и выходной.

1.3. **Случай с ограничениями на класс решающих правил.** В ряде работ [4, 5] предлагается строить кусочную аппроксимацию так, чтобы классификация производилась по одному набору показателей, а аппроксимация в каждом классе – по другому. Будем считать, что кроме пространства X есть ещё пространство $Z = \mathbf{R}^s$, в котором производится классификация объектов. Дальнейшие рассуждения не изменятся, если часть показателей в пространствах X и Z будут одни и те же.

Считается, что каждый из n объектов исходной выборки, описывается $k + s + 1$ параметром, т.е. вектором $(y_t, x_t, z_t) = \left(y_t, x_t^{(1)}, ..., x_t^{(k)}, z_t^{(1)}, ..., z_t^{(s)}\right) \in \mathbf{R}^{k+s+1}$.

Самый используемый критерий качества классификации – критерий средневзвешенной дисперсии, который для пространства Z запишем как функционал $J(\alpha_1, ..., \alpha_r, H(z)) = \sum\limits_{i=1}^{r} \sum\limits_{t=1}^{n} (z_t - \alpha_i)^2 \varphi(h_i(z_t))$, зависящий от эталонов классов $\alpha_1, ..., \alpha_r$ и вектор-функции принадлежностей $H(z)$. Считается, что эталоны классов могут быть произвольными точками пространства Z, вектор-функция $H(z)$ удовлетворяет условиям (3), а функция φ равна либо φ_1, либо φ_2, либо φ_3.

Минимизация этого функционала производится как по классификации $H(z)$, так и по набору эталонов классов $A = (\alpha_1, ..., \alpha_r)$. В оптимальном случае эталон i-го класса записывается в виде $\alpha_i = \frac{\sum\limits_{t=1}^{n} z_t \varphi(h_i(z_t))}{\sum\limits_{t=1}^{n} \varphi(h_i(z_t))}$, т.е. совпадает с центром класса.

Если $\varphi(t) = \varphi_1(t)$, то итерационный алгоритм минимизации этого функционала совпадает с известным алгоритмом ISODATA [2]. Если $\varphi(t) = \varphi_2(t)$, то – с размытым вариантом ISODATA. Если $\varphi(t) = \varphi_3(t)$, то оптимальная классификация дает размытые границы.

Если фиксирован набор эталонов классов $A = (\alpha_1, ..., \alpha_r)$, то эталонная классификация $H^A(z) = \left(h_1^A(z), ..., h_r^A(z)\right)$ для каждого из трёх вариантов функции $\varphi_j(h)$ однозначно определяется следующим образом.

1. Для $\varphi_1(h)$ $h_i^A(x) = \begin{cases} 1, & i = \arg\min\limits_{j=1,...,r}(z - \alpha_j) \\ 0, & i \neq \arg\min\limits_{j=1,...,r}(z - \alpha_j) \end{cases}$.

2. Для $\varphi_2(h)$ $h_i^A(x) = \frac{(z-\alpha_i)^{\frac{2}{1-t}}}{\left[\sum\limits_{j=1}^{r}(z-\alpha_j)^{\frac{2}{1-t}}\right]}$.

3. Случай функции $\varphi_3(h)$.

Для каждого объекта и каждого центра класса подсчитываются числа

$$v_i^{(1)} = t - (z - \alpha_i)^2 \sqrt{\frac{(r-1)t^2 + (t-1)^2}{\sum\limits_{j=1}^{r}(z - \alpha_j)^4}}.$$

Обозначим через $sign(v_i^{(1)}) = \begin{cases} 1, & v_i^{(1)} > 0 \\ 0, & v_i^{(1)} \leq 0 \end{cases}$.

Пусть r_- – число классов, для которых $sign(v_i^{(1)}) = 1$.Определим числа

$$v_i^{(1)} = sign(v_i^{(1)}) \left[t - (z - \alpha_i)^2 \sqrt{\frac{(r_- - 1)t^2 + (t-1)^2}{\sum\limits_{j=1}^{r} sign(v_i^{(1)}) (z - \alpha_j)^4}} \right].$$

По ним определяем функции принадлежностей

$$h_i^A(x) = t - \sqrt{t^2 - (2t-1)v_i^{(1)}}.$$

Таким образом, для любого набора из r векторов пространства Z можно определить эталонную классификацию$H^A(z) = \left(h_1^A(z), ..., h_r^A(z)\right)$. При решении задачи аппроксимации ограничимся множеством эталонных классификаций пространства Z.

Постановка задачи аппроксимации**:** минимизировать функционал кусочной аппроксимации (1), (2) или (3) при условии, что классификация $H(x)$ является эталонной в пространстве Z. Свободными параметрами, по которым необходимо оптимизировать функционал качества аппроксимации, являются: во-первых, набор эталонов классов $A = (\alpha_1, ..., \alpha_r)$, задающих эталонную классификацию; а во-вторых, коэффициенты линейных моделей каждого из классов $c_i, d_i, i = 1, ..., r$. Всего $(l + k + 1)\ r$ параметров. Перепишем функционалы (1), (2) и (3) в виде:

$$I_1' (A; c_i, d_i, i = 1, ..., r) = \sum_{i=1}^{r} \sum_{t=1}^{n} [y_t - ((c_i, x_t) + d_i)]^2\, h_i^A (z_t), \tag{5}$$

$$I_2' (A; c_i, d_i, i = 1, ..., r) = \sum_{t=1}^{n} \left[y_t - \sum_{i=1}^{r} ((c_i, x_t) + d_i)\ h_i^A (z_t) \right]^2, \tag{6}$$

$$I_3' (A; c_i, d_i, i = 1, ..., r) = \sum_{i=1}^{r} \sum_{t=1}^{n} [y_t - ((c_i, x_t) + d_i)]^2\, \varphi\left(h_i^A (z_t)\right). \tag{7}$$

Если в эталонной классификации $\varphi (t) = \varphi_2 (t)$ или $\varphi (t) = \varphi_3 (t)$, то функционалы (4)–(6) дифференцируемы по своим свободным параметрам и для нахождения их локальных экстремумов можно применять градиентные процедуры.

Недостатком алгоритмов локальной оптимизации является сильная зависимость результата от начальных условий работы алгоритма. Поэтому актуальным является разработка методов глобальной оптимизации.

2. Алгоритм построения кусочной аппроксимации для конечного множества эталонов

Заметим, что если зафиксировать набор эталонов классов $A = (\alpha_1, ..., \alpha_r)$, то однозначно по методу наименьших квадратов находятся коэффициенты линейных моделей классов $c_i, d_i,\ i = 1, ..., r$, минимизирующие один из функционалов (4) – (6). Таким образом, если можно перебрать все возможные наборы эталонов классов, то можно найти глобальный минимум выбранного функционала.

Пусть в пространстве Z выделено некоторое конечное множество точек $Z_p = \{\beta_1, ..., \beta_p\}$. Будем считать, что эталоны классов можно выбирать только из множества Z_p. Число вариантов выбора различных эталонов будет равно p^r. Так как в прикладных задачах число классов в кусочной аппроксимации редко бывает больше пяти-шести, а число точек в Z_p для подавляющего числа прикладных задач порядка 100, то такой перебор вполне можно делать на современных ПЭВМ.

В качестве множества Z_p можно взять, например, реализацию в пространстве Z исходной выборки объектов. В данном случае $Z_n = \{z_1, ..., z_n\}$. В качестве множества Z_p можно взять также достаточно разреженную решетку в пространстве Z. Такой вариант хорошо использовать в качестве начальных условий для градиентного алгоритма построения кусочной аппроксимации без ограничения на набор эталонов.

Следует отдельно выделить случай, когда классификационное пространство Zсоставляет один показатель.

2.1. Кусочная аппроксимация для одномерного классифицирующего пространства. Общая постановка задачи классификации не учитывает специфику одномерной классификации, а именно – упорядоченность точек числовой прямой. В результате в оптимальной классификации все точки (за исключением эталонов классов

$\alpha_1, ..., \alpha_r$) принадлежат с некоторым весом всем классам одновременно. Более того, функция принадлежности i-го класса $h_i(x)$ не унимодальная и имеет следующую структуру: на отрезке $[\alpha_{i-1}, \alpha_i]$ $h_i(x)$ возрастает от 0 до 1, на отрезке $[\alpha_i, \alpha_{i+1}]$ она убывает от 1 до 0, на отрезке $[\alpha_{j-1}, \alpha_j]$ $(j \neq i, i+1)$ она возрастает от 0 до некоторой величины $b < 1$ и затем опять убывает до 0, при $x \to \infty$ $h_i(x) \to {}^1\!/_r$.

В случае одномерной классификации естественно предполагать, что перекрываться могут лишь соседние классы.

В итоге оказывается, что, α_i – не только эталон класса, но и граница между (i -1)-м и (i+ 1)-м классами. В соответствии с этим на эталонную классификацию $H^A(z) = \left(h_1^A(z), ..., h_r^A(z)\right)$ наложим следующие дополнительные ограничения (предполагается, что $\alpha_1 \leq \alpha_2 \leq ... \leq \alpha_r$):

- если $z \leq \alpha_1$, то $h_1^A(z) = 1$, $h_j^A(z) = 0$, $j \neq 1$, т.е. на этом луче все точки однозначно относятся к первому классу;
- если $\alpha_{i-1} \leq z \leq \alpha_i$, то $h_{i-1}^A(z) \geq 0$, $h_i^A(z) \geq 0$, $h_j^A(z) = 0$, $j \neq i-1$, $j \neq i$, т.е. на этом отрезке ненулевые веса могут иметь лишь i-й класс и (i-1)-й;
- если $\alpha_r \leq z$, то $h_r^A(z) = 1$, $h_j^A(z) = 0$, $j \neq r$, на этом луче все точки однозначно относятся к последнему классу.

Заметим, что функционалы (4) и (6) отличаются лишь наличием функции φ. Другими словами (4) – это (6), в котором $\varphi(h) = h$. Далее рассматривается алгоритм для (6).

Эталоны классов разбивают числовую прямую на (r+ 1) промежутков, в каждом из которых могут применяться не более двух локальных линейных моделей кусочной аппроксимации. В соответствии с этим перепишем функционал (6) в следующем виде:

$$I_3''(A; c_i, d_i, i = 1, ..., r) = \sum_{z_t \leq \alpha_1} (y_t - (c_1 x_t + d_1))^2 \quad + \sum_{\alpha_r \leq z_t} (y_t - (c_r x_t + d_r))^2 \quad +$$

$$+ \sum_{i=2}^{r} \quad \sum_{\alpha_{i-1} \leq z_t \leq \alpha_i} \left[(y_t - (c_{i-1} x_t + d_{i-1}))^2 \varphi\left(h_{i-1}^A(z_t)\right) + (y_t - (c_i x_t + d_i))^2 \varphi\left(h_i^A(z_t)\right)\right].$$

Обозначим $\Delta(i, \alpha, \beta, c, d) = \sum\limits_{\alpha \leq z_t \leq \beta} \left[(y_t - (c x_t + d))^2 \varphi\left(h_i^A(z_t)\right)\right]$., тогда

$$I_3''(A; c_i, d_i, i = 1, ..., r) = \Delta(1, \alpha_0, \alpha_1, c_1, d_1) \ + \sum_{i=2}^{r} \Delta(i-1, \alpha_{i-1}, \alpha_i, c_{i-1}, d_{i-1}) +$$
$$+ \sum_{i=2}^{r} \Delta(i, \alpha_{i-1}, \alpha_i, c_i, d_i) + \Delta(r, \alpha_r, \alpha_{r+1}, c_, d_1);$$

$$I_3''(A; c_i, d_i, i = 1, ..., r) = \sum_{i=1}^{r} (\Delta(i, \alpha_{i-1}, \alpha_i, c_i, d_i) + \Delta(i, \alpha_i, \alpha_{i+1}, c_i, d_i)).$$

Здесь предполагается, что $\alpha_0 = -\infty$, а $\alpha_{r+1} = +\infty$.

Если известны α_{i-1}, α_i и α_{i+1}, то однозначно известна функция $h_i(x)$, следовательно, по ней однозначно рассчитываются коэффициенты c_i и d_i.

Пусть $S_i(\alpha_{i-1}, \alpha_i, \alpha_{i+1}) = (\Delta(i, \alpha_{i-1}, \alpha_i, c_i, d_i) + \Delta(i, \alpha_i, \alpha_{i+1}, c_i, d_i))$, тогда

$I_3''(A; c_i, d_i, i = 1, ..., r) = \sum_{i=1}^{r} S_i(\alpha_{i-1}, \alpha_i, \alpha_{i+1})$. Для $i = r-1, ..., 1$ строим функции

$$F_r(\alpha_{r-1},\alpha_r) = S_r(\alpha_{r-1}, \alpha_r, \alpha_{r+1}), \quad F_i(\alpha_{i-1},\alpha_i) = \min_{\alpha_{i+1}} [S_i(\alpha_{i-1}, \alpha_i, \alpha_{i+1}) + F_{i+1}(\alpha_i,\alpha_{i+1})].$$

Последнее выражение – рекуррентное уравнение динамического программирования Беллмана. Решая задачу $F_0(\alpha_0) = \min_{\alpha_1} [S_1(\alpha_0, \alpha_1, \alpha_2) + F_1(\alpha_0,\alpha_1)]$, получим оптимальное значение α_1, по нему – значение α_2 и т.д. до получения оптимального значения α_r.

2.2. **Кусочная аппроксимация с классификацией по выходному параметру.** В задаче кусочной аппроксимации часто бывает полезно результирующую классификацию проектировать не на пространство X, а на y [1]. Тогда классы можно интерпретировать как режимы работы, которые обеспечивают соответствующий уровень y. Для этого достаточно рассмотреть случай $z = y$ и глобальный экстремум функционалов (1) и (3) находится с помощью процедуры динамического программирования.

Заключение

В работе рассмотрены алгоритмы решения задачи кусочно-линейной аппроксимации сложной зависимости с использованием вариационного подхода к задачам классификационного анализа данных. Даются постановки этой задачи, как для чёткой, так и для размытой классификаций. Рассматривается два способа нахождения глобального экстремума функционала – для случая конечного множества эталонов и для одномерного классификационного пространства. Для последнего случая построена рекуррентная схема Беллмана (динамическое программирование), реализация которой и обеспечивает получение глобального экстремума. Рассмотрен важный для приложений случай, когда одномерное классификационное пространство – выходной параметр.

Список литературы

1. *Райбман Н.С., Дорофеюк А.А., Касавин А.Д.* Идентификация технологических объектов методами кусочной аппроксимации. М.: ИПУ, 1977. – 70 с.
2. *Бауман Е.В.,Дорофеюк А.А.* Классификационный анализ данных. В сб.: «Избранные труды Международной конференции по проблемам управления. Том 1». М.: СИНТЕГ, 1999.
3. *Bauman E.V.* Variational Methods of Piecewise Approximation. Proc. of 8-th Conference on Analysis and Optimization Systems. (Antibes, France, 1988), INRIA, 1988.
4. *Бауман Е.В.,Дорофеюк А.А., Чернявский А.Л.* Методы структурной обработки эмпирических данных. Измерение, контроль, автоматизация. 1985, № 3. 5.
5. *Бауман Е.В., Блудян Н.О.* Методы нахождения глобальных экстремумов функционалов в задаче классификационного анализа данных. Труды ИПУ РАН, т. XIII. М.: ИПУ, 2001, С. 129–136.

Статья поступила в редакцию 28.04.2008