

УДК 518:621.391

СИНТЕЗ ИЕРАРХИЧЕСКИХ МНОГОФАКТОРНЫХ  
МОДЕЛЕЙ И ИХ АДАПТАЦИЯ

В.И.Котюков, А.Е.Буторин

Рассматривается задача оптимального построения различного типа иерархических (в частности, кусочно-простых) моделей прогнозирования значений случайной величины  $Y$  на основе "замеряемых" значений показателей  $X = (X_1, X_2, \dots, X_n)$ . Отсутствие эффективных численных методов построения сложных (нелинейных) многофакторных математических моделей  $Y^* = g(X)$ , а также ряд других причин вызвали в последнее время повышенный интерес к использованию принципа иерархичности. При этом мы сталкиваемся с вопросами кусочно-линейной аппроксимации дискриминантной и регрессионной функций, организацией различного рода "коллективов" ("комитетов") простых моделей прогноза и другими задачами.

Иерархические модели, обладая простой интерпретацией (в отличие от так называемых непараметрических статистических моделей), позволяют получать достаточно адекватное приближение к оптимальной  $Y_0^* = g_0(X)$  модели. Кроме того, мы имеем возможность эффективно решать задачи оптимального управления и планирования на основе уже построенных иерархических моделей.

Построение модели  $Y^* = g(X)$  обычно основано на анализе статистического материала (выборки) в  $M$  наблюдений (реализаций), для каждого из которых известно как значение вектора исходных показателей  $x = (x_1, \dots, x_n)$ , так и значение результирующего фактора  $y$ . Выборка  $\{(y_i, x_i) ; i = \overline{1, M}\}$  получена в соответствии с некоторой неизвестной нам функцией распределения  $F(y, x)$  в многомерном пространстве  $\{Y, X\}$ .

Если  $Y$  - непрерывная случайная величина (количественный показатель, шкала отношений), то мы имеем дело с задачей регрессионного анализа, если же  $Y$  замерен в шкале наименований (классов,

типов), то - с задачей дискриминантного анализа (распознавания образов).

Основное внимание здесь будет уделено построению кусочно-простой модели прогноза, сущность которой заключается в том, что все множество возможных реализаций  $R$  подпространства  $X$  разбивается некоторым способом на такие подмножества, в пределах которых простые модели прогноза будут уже вполне адекватны. Оптимизация построения кусочно-простой модели на основе анализа выборки осуществляется с помощью разработанного метода "фиксированных элементов пространства разбиений" [1], представляющего собой единый подход к синтезу  $Y^* = g(X)$  как для задач регрессионного, так и дискриминантного анализа.

### §1. Формальная постановка задачи

Допустим, что все множество возможных реализаций  $R$  разбито некоторым образом на  $k$  непересекающихся подмножеств (областей)  $\{O_1, \dots, O_k\}$ . Тогда прогноз по кусочно-простой модели имеет вид

$$y^* = \sum_{v=1}^k \psi(x, \beta_v) g(x, \alpha_v),$$

где

$$\psi(x, \beta_v) = \begin{cases} 1, & \text{если } x \in O_v; \\ 0, & \text{если } x \notin O_v; \end{cases} \quad x = (x_1, \dots, x_n); \quad (I)$$

$$\bigcup_{v=1}^k O_v = R; \quad O_v \cap O_e = \emptyset \quad \text{при } v \neq e;$$

$g(x, \alpha_v)$  - некоторая "простая" параметрическая модель прогноза для области  $O_v$ . Такой моделью может быть, например, гиперплоскость  $g(x, \alpha_v) = \sum_{j=1}^n \alpha_{vj} x_j + \alpha_{v0}$  в задаче кусочно-линейной аппроксимации функции регрессии.

Оптимальный синтез кусочно-простой модели предусматривает выбор таких значений величины  $k$  и векторов параметров  $\{\alpha_v, \beta_v\}$ , при которых достигает минимума некоторая оценка ошибки прогноза  $\delta$  по данной модели.

При такой постановке данную задачу можно рассматривать и как задачу таксономии при известном внешнем критерии качества решения [2].

Считаем, что оценка ошибки прогноза обладает аддитивной структурой

$$\delta = \sum_{v=1}^k \delta_v \quad (2)$$

где  $\delta_v$  - соответствует оценке ошибки прогноза по  $g(X, \alpha_v)$  для реализаций  $\{(y, x) : x \in O_v\}$  и удовлетворяет соотношению

$$\delta_v \geq \tilde{\delta}_v / N \quad ; \quad (3)$$

здесь  $\tilde{\delta}_v$  - суммарная ошибка прогноза на  $N_v$  выборочных реализациях  $\{(y_i, x_i) : x_i \in O_v\}$  по модели  $g(X, \alpha_v)$ .

Например, для задачи регрессионного анализа в качестве оценки  $\delta_v$  для модели  $g(X, \alpha_v) = \sum_{j=1}^{n_v} \alpha_{vj} X_j + \alpha_{v0}$  (при  $n_v \leq n$ ) может быть взята величина

$$\delta_v = \left( \frac{N_v}{N} \right) \frac{1}{N_v - n_v - 1} \tilde{\delta}_v \quad .$$

где

$$\tilde{\delta}_v = \sum_{i=1}^{N_v} (y_i - g(x_i, \alpha_v))^2 \quad .$$

Соотношениям (2) и (3) удовлетворяют некоторые критерии "эмпирического риска" [3], а также ряд других оценок.

Конкретный вид  $\delta$  нас интересовать не будет. Заметим, что минимизация  $\delta$  позволяет выбрать и наилучшее значение величины  $k$ .

Предполагается также, что существует некоторый алгоритм  $A$ , позволяющий на элементах (реализациях), принадлежащих произвольно выделенному подмножеству  $O_v$ , построить свою оптимальную по  $\delta_v$  "простую" модель  $g(X, \alpha_v)$ . Для ранее рассмотренной задачи регрессионного анализа таким методом является метод решения нормальных уравнений Гаусса.

Предметом нашего рассмотрения является оптимизационный метод разбиения  $R$  на области  $\{O_v\}$ . Характеризация областей  $\{O_v\}$  осуществляется с помощью построения разделяющих их поверхностей в пространстве  $X$ .

Считаем, заданной своими параметрами систему разделяющих "границ"  $\Gamma = \{\Gamma^1(X), \dots, \Gamma^e(X)\}$ , относительно каждой из которых на конкретной реализации  $x$  выполняется либо  $\Gamma^e(x) = 0$ , либо  $\Gamma^e(x) = 1$ ,  $e = \overline{1, s}$  [1]. В простейшем и нами практически используемом

варианте системе  $\Gamma$  соответствует список  $\{z^1, \dots, z^s\}$  некоторых значений признаков  $\{X_j\}$ . Пусть  $z^e$  - некоторое фиксированное значение  $X_j$ . Тогда если  $X_j$  - количественный показатель или шкала порядка, то  $\Gamma^e(x) = 0$  при  $x_j \leq z^e$  и  $\Gamma^e(x) = 1$  при  $x_j > z^e$ . Если же  $X_j$  - шкала наименований, то  $\Gamma^e(x) = 0$  при  $x_j = z^e$  и  $\Gamma^e(x) = 1$  при  $x_j \neq z^e$ .

Таким образом, система  $\Gamma = \{\Gamma^1(X), \dots, \Gamma^s(X)\}$  (или просто  $\{\Gamma^1, \dots, \Gamma^s\}$ ) есть совокупность некоторых булевых признаков - "границ", фиксированных до синтеза разбиения  $\{O_v\}$ . Области  $\{O_v\}$  разделяются поверхностями, сформированными на основе элементов  $\{\Gamma^e\}$  согласно нижеописанной схеме.

Список  $\Gamma$  фиксирует список подмножеств  $P = \{P^1, \dots, P^s\}$  ( $s = 2s$ ), где  $P^e = \{x: \Gamma^e(x) = 0\}$ ;  $P^{e+1} = \{x: \Gamma^e(x) = 1\}$ ,  $t = 2e-1$ .

Перейдем к определению допустимого разбиения  $\{O_1, \dots, O_k\}$ .

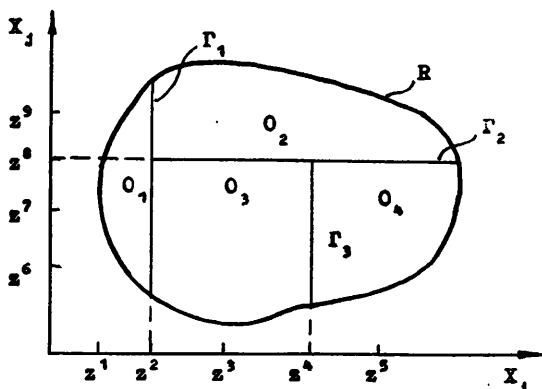
Последовательность  $(k-1)$  границ  $(\Gamma_1, \dots, \Gamma_{k-1})$ , выбранных из списка  $\Gamma$  (обычно  $k \ll s$ ) с тем или иным возможным значением (0 или 1), а стало быть, и выбранная из списка  $P$  последовательность соответствующих подмножеств  $(P_1, \dots, P_{k-1})$  определяет систему  $k$  областей  $\{O_v\}$  следующим образом:

$$\left. \begin{aligned} x \in P_1 &\Rightarrow x \in O_1; \\ x \bar{\in} O_1 \wedge x \in P_2 &\Rightarrow x \in O_2; \\ \dots &\dots \dots \\ x \bar{\in} O_1 \wedge \dots \wedge x \bar{\in} O_{k-2} \wedge x \in P_{k-1} &\Rightarrow x \in O_{k-1}; \\ x \bar{\in} O_1 \wedge \dots \wedge x \bar{\in} O_{k-1} &\Rightarrow x \in O_k. \end{aligned} \right\} \quad (4)$$

Здесь, например, вторую формулу следует понимать так: если  $x \bar{\in} O_1$  и  $x \in P_2$ , то в этом случае  $x \in O_2$ . Иными словами, область  $O_v$  "формируется" границей  $\Gamma_v$  (с соответствующим значением) при уже определенных подмножествах  $(O_1, \dots, O_{v-1})$ .

Разбиение  $(O_1, \dots, O_k)$ , удовлетворяющее системе (4), будем называть основным допустимым разбиением. На рисунке приведен пример такого разбиения.

Задача заключается в таком разбиении множества  $R$  в пространстве  $X$  на  $k$  непересекающихся подмножеств  $\{O_v\}$ , удовлетворяющих (4), при котором система соответствующих оптимальных "простых" моделей  $\{g(X, \alpha_v)\}$  будет обеспечивать минимум (2). Условия (4) сводят задачу синтеза оптимального разбиения  $\{O_v\}$  к задаче выбора оптимальной последовательности  $(k-1)$  границ  $(\Gamma_1, \dots, \Gamma_{k-1})$  из



списка  $\Gamma$  и соответствует -  
 вующих им возможных  
 значений (0 или 1), или,  
 что то же самое, к за-  
 даче выбора оптималь-  
 ной последовательности  
 $(P_1, \dots, P_{k-1})_0$  из спи-  
 ска  $P$ .

## §2. Оптимизация решения

В основу предлагаемого ниже алгоритма синтеза оптимальных областей  $\{O_1, \dots, O_k\}$  положен метод "ветвей и границ" [4].

Для простоты рассмотрения пусть значение  $k$  заранее фиксировано.

Процесс выбора  $(P_1, \dots, P_{k-1})_0$  из списка  $P$  имеет "ветвящуюся" структуру: на первом шаге рассматриваются размещения из  $m$  исходных подмножеств  $\{P^1, \dots, P^m\}$  по одному  $\{(P^i)\}$ , на втором шаге - по два  $\{(P^i, P^j)\}$  и так далее. При этом мы для каждой "ветви"  $(P^i, P^j, \dots)$  и для каждой из областей  $\{O_v\}$ , получаемой согласно (4), строим свою оптимальную "простую" модель  $g(X, \alpha_v)$  и определяем оценку ошибки прогноза для нее  $\delta_v$ .

Индекс  $p$  будет использоваться в случае необходимости выделить некоторую ветвь длины  $e$   $(P_1, \dots, P_e)_p$ , соответствующие ей области  $\{O_{1p}, \dots, O_{ep}\}$ , модели  $\{g(X, \alpha_{vp})\}$  и соответствующие ошибки  $\{\delta_{1p}, \dots, \delta_{ep}\}$ . Пусть  $e < (k-1)$ . Данной ветви соответствует оценка ошибки

$$\delta_p = \sum_{v=1}^e \delta_{vp}.$$

Однако при переходе от  $e$ -го шага к  $(e+1)$ -у не все ветви длины  $e$  служат основой для образования ветвей длины  $(e+1)$ . Дело в том, что для любой ветви  $(P_1, \dots, P_e)_p$  оказывается возможным определить заведомо наилучший  $(\delta_p \min)$  и наихудший из приемлемых  $(\delta_p \max)$  варианты ее продолжения.

Обозначим через  $\delta_p^*$  - оценку ошибки для "простой" модели  $g(X, \alpha_p^*)$ , построенной на выборочных реализациях, принадлежащих всей области "оставшейся" от  $p$ -й ветви  $O_p^* = R \setminus \bigcup_{v=1}^e O_{vp}$ .

Пусть существует такая уже "синтезированная"  $q$ -я ветвь, что для ряда порожденных ею моделей  $\{g(X, \alpha_{jq})\}$ , общая область действия которых  $\bigcup_j O_{jq} \supset O_p^*$  (либо может быть искусственно распространена на  $O_p^*$ ), суммарная оценка их ошибок на реализациях  $\{(y_i, x_i) : x_i \in O_p^*\}$  равна  $\sum_j \delta_{jq}^{*p}$ . Для алгоритма с фиксированным  $k$  число таких моделей должно быть не более  $(k - e)$ . Тогда

$$\delta_p \max = \delta_p + \min \left\{ \delta_p^*, \min_{q-j} \left\{ \sum \delta_{jq}^{*p} \right\} \right\}. \quad (5)$$

Определим величину  $\delta_p \min$ . Пусть на данном шаге алгоритма имеется совокупность  $R_\mu$  непересекающихся областей  $\{O_{jt}\} - R = \bigcup_{(jt)} O_{jt} \subset O_p^*$ . Известно также, что суммарная ошибка  $\tilde{\delta}_\mu = \sum_{(jt)} \tilde{\delta}_{jt}$  (см. пояснения к (3)) соответствующих моделей  $\{g(X, \alpha_{jt})\}$  на выборочных реализациях  $\{(y_i, x_i) : x_i \in R_\mu\}$  не может быть уменьшена на  $R_\mu$  за счет построения других моделей. Тогда

$$\delta_p \min = \begin{cases} \delta_p + \max_{\mu} \{ \tilde{\delta}_\mu / N \}, & \text{если } \exists \mu; \\ \delta_p + 0 = \delta_p, & \text{если } \bar{\exists} \mu. \end{cases} \quad (6)$$

Наиболее же просто вычислим оценки [I]

$$\{ \delta_p \min = \delta_p; \quad \delta_p \max = \delta_p + \delta_p^* \}. \quad (7)$$

Очевидно, что любое "продолжение"  $p$ -й ветви, дающее значение  $\delta > \delta_p \max$ , а такое возможно, является неприемлемым и его "синтезировать" не следует.

Определим величину "рекорда" по всем построенным на данном шаге ветвям

$$r = \min_p \{ \delta_p \max \}. \quad (8)$$

Ветви, у которых  $\delta_p \min \geq r$ , в дальнейшем не рассматриваются и не служат основой для образования ветвей большей длины. Полученная на  $(k-1)$ -м шаге ветвь, для которой величина  $\delta$  минимальна, и является искомой  $(P_1, \dots, P_{k-1})_0$ .

Анализ приведенного метода позволяет сделать вывод о возможности его использования и для случая заранее не фиксированного значения  $k$ .

Метод позволяет учитывать априорное ограничение  $N_v \geq N_0$ , где  $N_0$  - минимально допустимое число выборочных реализаций для любой области  $Q_v$ . Знание  $N_0$  позволяет значительно сокращать "переборы" возможных ветвей при поиске  $(P_1, \dots, P_{k-1})_0$ .

Программная реализация для ЭВМ данного метода может быть осуществлена согласно различным принципам: "в ширину" (как описано), "в глубину" (когда на каждом шаге рассматривается очередная ветвь  $(P_1^i, P_2^j, \dots, P_{k-1}^q)$  полной длины) и на основе разумного сочетания двух вышеупомянутых принципов.

Однако при больших величинах  $n$ ,  $M$  и  $m$  программная реализация на ЭВМ рассмотренного оптимального алгоритма может встретить затруднения как по быстродействию, так и по требуемому объему оперативной памяти. Ниже будет изложен приближенный метод - метод "гарантированного включения" [5].

Алгоритмическая структура метода "гарантированного включения" проста. Предполагается, что после первого шага в памяти "хранится"  $m$  ветвей, каждая из которых в качестве своей единственной компоненты содержит некоторое  $P^i$  ( $i = \overline{1, m}$ ). Пусть после  $e$ -го шага в памяти также "хранится"  $m$  ветвей длины  $e$ . Тогда на  $(e+1)$ -м шаге выполняются следующие операции. Элемент  $P^i$  мы пытаемся "присоединить" ко всем из  $m$  ветвей, полученных на  $e$ -м шаге и еще не содержащих  $P^i$ . Из полученных ветвей длины  $(e+1)$ , содержащих  $P^i$  в качестве последней компоненты, запоминается лишь одна - наилучшая. Такие же операции выполняются и для всех других  $P^i$  ( $i = \overline{1, m}$ ). В результате  $(e+1)$ -го шага в памяти будет "храниться" также  $m$  ветвей, причем каждая  $P^i$  будет включена по крайней мере в одну из них (принцип гарантированного включения) и т.д. Лучшая ветвь, полученная на  $(k-1)$ -м шаге, является искомой.

Алгоритм "гарантированного включения" можно дополнять принципами "отсечения" оптимального метода.

В общем случае данный алгоритм гарантирует получение  $(P_1, \dots, P_{k-1})_0$  лишь при  $n=1$  ( $k$  - любой) или при  $k \leq 3$  ( $n$  - любое).

Решение на ЭВМ модельных оптимизационных задач позволяет сделать вывод о высокой точности метода "гарантированного включения" в смысле большой вероятности малого отклонения от оптимального решения.

### §3. Оптимизация сложных разбиений

Под сложным разбиением  $\{O_1, \dots, O_k\}$  будем понимать такое, у которого каждая граница  $\Gamma_i$  из выбранного списка  $\{\Gamma_1, \dots, \Gamma_{k-1}\}$  отделяет одну совокупность областей  $\{O_v\}$  от другой  $\{O_g\}$ . Сложное разбиение можно рассматривать как конечную композицию разбиений типа (4) (любая область  $O_v$  в разбиении типа (4) может не сопоставляться с  $g(X, \alpha_v)$ , а, в свою очередь, допускать разбиение типа (4)).

Для синтеза сложного разбиения модифицируем ранее описанный оптимальный (для основного допустимого разбиения) метод. Пусть "ветвящийся" процесс "дробления"  $R$  имеет следующую структуру. На первом шаге каждая из исходных границ  $\Gamma^i$  порождает  $e$  различных разбиений, содержащих по две соответствующих области, для каждой из которых строится модель  $g(X, \alpha_v)$  и определяется оценка ошибки  $\delta_v$ . На  $e$ -м шаге каждое из полученных на  $(e-1)$ -м шаге разбиений  $\{\{O_{1q}, \dots, O_{eq}\}\}$  с помощью границы  $\Gamma^i$  ( $i = \overline{1, e}$ ) порождает  $e$  новых разбиений, состоящих из  $(e+1)$  компонент каждая, путем "дробления" с помощью  $\Gamma^i$  каждой области  $O_{jq}$  на две ( $j = \overline{1, e}$ ) и т.д.

Для каждого  $q$ -го разбиения  $\{O_{1q}, \dots, O_{eq}\}$  возможно определить наилучший ( $\delta_{q \max}$ ) и наилучший ( $\delta_{q \min}$ ) варианты его дальнейшего "дробления". Определим для каждой из областей  $O_{vq}$  аналогичные значения  $\{\delta_{vq \max}, \delta_{vq \min}\}$  по методике, приведенной для области  $O_p^*$  в предыдущем параграфе (см. соответствующие компоненты формул (5) и (6)). Тогда

$$\delta_q \max = \sum_{v=1}^e \delta_{vq \max}, \quad \delta_q \min = \sum_{v=1}^e \delta_{vq \min}. \quad (9)$$

Принцип "отсечения" на каждом шаге неперспективных разбиений на основе значений (9) аналогичен принципу для основных допустимых разбиений.

Очевидно, что трудоемкость оптимизации сложного разбиения намного больше, чем основного допустимого.

Рассмотрим возможность приближенной оптимизации сложных разбиений. Пусть на первом этапе решения задачи реализуется ранее описанный алгоритм "гарантированного включения" с учетом следующего требования: если на  $(e+1)$ -м шаге принимается окончательное решение о присоединении компоненты  $P^i$  к некоторой ветви  $P_1, \dots, P_e$ , то и "двойственная" компонента (порожденная той же границей  $\Gamma^i$ )  $P^{i+1}$  также присоединяется к этой ветви.



На втором этапе осуществляется анализ всех полученных допустимых разбиений с целью возможного образования из них сложных разбиений. Анализ проводится в обратном направлении (от  $(e+1)$ -го шага к  $e$ -у) согласно следующему принципу. Пусть для  $q$ -го разбиения имеет место  $\{O_{1q}, \dots, O_{eq}, O_{(e+1)q}, R_q\}$  и для  $p$ -го -  $\{O_{1p}, \dots, O_{ep}, O_{(e+1)p}, R_p\}$ , где  $R_q$  и  $R_p$  - некоторые совокупности областей с суммарными оценками ошибок  $\delta_q^i$  и  $\delta_p^i$  соответственно; подразбиения  $\{O_{1q}, \dots, O_{(e+1)q}\}$  и  $\{O_{1p}, \dots, O_{(e+1)p}\}$  имеют структуру допустимых разбиений;  $O_{vq} = O_{vp}$  для  $v = \overline{1, e}$ ;  $O_{(e+1)q} = R_p$  и  $O_{(e+1)p} = R_q$ . Пусть также  $\delta_{(e+1)q} > \delta_p^i$ . Тогда исходное  $q$ -е разбиение заменяется на сложное разбиение  $\{O_{1q}, \dots, O_{eq}, R_q^i\}$ , где  $R_q^i = R_q \cup R_p$ . Такой анализ нужно осуществлять для всех подветвей, полученных на  $e$ -м шаге первого этапа. Возможность анализа существует в силу ранее упомянутого дополнительного требования к методу "гарантированного включения".

Разбиение, полученное после анализа всех подветвей длины  $l$  и имеющее наименьшее значение  $\delta$ , является искомым.

#### §4. Возможности метода "фиксированных элементов пространства разбиений"

Приведем сравнительный анализ рассматриваемого метода и других методов построения кусочно-простых моделей прогнозирования.

Недостатком метода "фиксированных элементов пространства разбиений" является отсутствие возможности построения сложной поверхности, разделяющей каждую пару областей, например  $(O_v, O_e)$ , в виде гиперплоскости в многомерном пространстве  $X$ . Однако он обладает рядом положительных качеств. Рассмотрим их.

1. Метод представляет собой единый подход к построению кусочно-простой модели как для задачи регрессионного, так и дискриминантного анализа (распознавания образов). С его помощью может решаться и задача кластерного анализа (таксономии). Специфика конкретной задачи отражается в формулировке критерия  $\delta$  и способе определения величин  $\{\delta_p^{\max}, \delta_p^{\min}\}$ .

2. Метод представляет собой единый подход к синтезу различного типа кусочно-простых моделей. На практике, однако, наибольшее распространение получили кусочно-постоянные, кусочно-линейные, кусочно-квадратичные и кусочно-логические модели. Заметим, что кусочно-логические модели являются частным случаем логических решающих функций [6], которым также присущи рассматриваемые здесь свойства.

3. В рамках метода можно предложить эффективные численные алгоритмы оптимизации кусочно-простой модели, обеспечивающие достижение глобального экстремума функционала качества  $\delta$ .

4. Границы, разделяющие области  $\{O_v\}$ , имеют простую смысловую интерпретацию для специалистов (они перпендикулярны исходным координатным осям  $\{X_j\}$ ), и области  $\{O_v\}$  попарно линейно разделимы.

5. Оптимальное разбиение  $\{O_1, \dots, O_k\}$  можно реализовать в пространстве разнотипных показателей  $\{X_j\}$ .

6. Описание разделяющих границ  $\{\Gamma_i\}$  содержит лишь статистически значимые (информативные) показатели из списка  $X$ .

7. "Синтезируя" разбиение  $\{O_v\}$  мы имеем возможность включить в каждую простую модель  $g(X, \alpha_v)$  также лишь статистически значимые и разнотипные признаки  $\{X_j\}$ . Алгоритм построения линейных регрессионных и дискриминантных моделей  $g(X, \alpha_v)$  от разнотипных показателей  $\{X_j\}$  описан в [7] и основан на пошагово-оптимальном числовом перекодировании классификационных признаков  $\{X_q\}$ .

8. В рамках метода возможно эффективное решение задачи структурной адаптации кусочно-простых моделей в процессе накопления новой информации  $\{(y_1, x_1)\}$ .

### §5. Адаптация кусочно-простой модели

При решении многих прикладных задач (например, при управлении технологическим процессом производства какой-либо продукции) желательно уметь проводить адаптацию модели прогноза при поступлении новой информации  $\{(y_1, x_1)\}$ , не выполняя заново полного расчета  $Y^* = g(X)$ . В вопросу параметрической адаптации кусочно-простых моделей (корректировка подмоделей  $\{g(X, \alpha_v)\}$  и границ  $\{\Gamma_v\}$ ) посвящена работа [5].

Наибольшую же трудность (в основном из-за сложных оптимизационных расчетов) в решении проблемы адаптации представляет выбор количества  $k$  областей разбиения  $\{O_v\}$  и соответствующих разделяющих границ  $\{\Gamma_v\}$ , то есть структурная адаптация.

Предлагаемый ниже алгоритм структурной адаптации кусочно-простой модели  $Y^* = g(X)$  также основан на использовании основных предпосылок рассматриваемого метода. Считаем, что по мере поступления данных  $\{(y_1, x_1)\}$  относительно каждой из исходных границ  $\Gamma^0$  рекуррентно вычисляются следующие статистики: число  $n_v^0$  реализаций  $(y_1, x_1)$ , у которых  $x_1 \in O_v \wedge \Gamma^0(x_1) = 0$ , суммар-

ная ошибка прогноза  $\delta_v^e$  этих  $N_v^e$  реализаций по  $g(X, \alpha_v)$  ( $v = \overline{1, k}$ ), суммарная ошибка прогноза  $\delta_{vq}$  на реализациях  $\{x_i \in O_v\}$  по модели  $g(X, \alpha_q)$  и ошибка  $\delta_{qv}$  на реализациях  $\{x_i \in O_q\}$  по модели  $g(X, \alpha_v)$  при условии, что области  $(O_v, O_q)$  - "соседние" (двойственные). Наличие этих характеристик  $\{N_v^e, \delta_v^e, \delta_{vq}, \delta_{qv}\}$ , полученных относительно исходных элементов  $\{\Gamma^1, \dots, \Gamma^e\}$ , позволяет на каждом шаге решать задачи структурной адаптации модели. Эти задачи решаются традиционными методами проверки статистических гипотез. При этом мы можем принимать решение об объединении двух "соседних" областей  $(O_v, O_q)$  в одну, если для подмоделей  $g(X, \alpha_v)$  и  $g(X, \alpha_q)$  имеет место (с заданной доверительной вероятностью  $\gamma$ ) соотношения  $\delta_{vq} \approx \delta_v$  и  $\delta_{qv} \approx \delta_q$ , либо принимать решения о дроблении области  $O_v$  на две, если, подмодель  $g(X, \alpha_v)$  не обеспечивает требуемой ошибки прогноза  $\delta_0$  с заданной доверительной вероятностью  $\gamma$ . Решение о дроблении  $O_v$  на две области принимается лишь тогда, когда выявится та новая разделяющая граница  $\Gamma^e$ , относительно которой имеем неадекватность модели  $g(X, \alpha_v)$  с вероятностью  $\gamma$  как для подмножества  $\{(y_i, x_i): x_i \in O_v \wedge \Gamma^e(x_i) = 0\}$ , так и  $\{(y_i, x_i): x_i \in O_v \wedge \Gamma^e(x_i) = 1\}$ . Заметим, что после проведения структурной адаптации некоторые статистики  $\{N_v^e, \delta_v^e, \delta_{vq}, \delta_{qv}\}$  просто пересчитываются на основе уже имеющейся информации, а некоторые приходится накапливать заново.

### §6. Коллективы кусочно-простых моделей

В последнее время находит широкое применение использование различного рода коллективов простых моделей для улучшения качества прогноза [8]. Здесь будет рассмотрен один из возможных вариантов коллектива кусочно-простых моделей прогноза.

Допустим имеется  $p$  различных кусочно-простых моделей  $\{O_{1e}, \dots, O_{ke}\}, (g(X, \alpha_{1e}), \dots, g(X, \alpha_{ke})); e = \overline{1, p}\}$  прогноза одной и той же случайной величины  $Y$ . Для простоты рассмотрения и не уменьшая его общности будем полагать, что у всех  $p$  моделей количество областей разбиения одинаково и равно  $k$ . Согласно (I) эти модели имеют представление

$$y_e^* = \sum_{v=1}^k \Psi(x, \beta_{ve}) \cdot g(X, \alpha_{ve}); \quad e = \overline{1, p}.$$

Пусть также еще на этапе обучения на основе исходной выборки определена матрица  $A$  размером  $(p \times k) \times (p \times k)$ , элементами кото -

рой являются ковариации между всеми подмоделями  $g(X, \alpha_{ve})$ ;  $v = \overline{1, k}$ ;  $e = \overline{1, p}$ . Заметим, что ковариация между  $g(X, \alpha_{ve})$  и  $g(X, \alpha_{jq})$ , как ковариация между некоторыми показателями, определяется лишь на подвыборке  $V = \{x: x \in O_{ve} \wedge x \in O_{jq}\}$ . Матрица  $A$  содержит лишь "достоверно" определенные ковариации. В задачах дискриминантного анализа подобные ковариационные матрицы определяются отдельно для каждого класса. Тогда процесс прогнозирования значения  $Y$  в точке  $x$  состоит из следующих четырех этапов:

- определяется область  $O_{ve}$  принадлежности точки  $x$  ( $x \in O_{ve}$ ) для каждой из  $p$  моделей ( $e = \overline{1, p}$ );
- по выбранным  $p$  областям  $\{O_{ve}\}$ , а стало быть, и  $p$  соответствующим подмоделям  $\{g(X, \alpha_{ve})\}$  выделяется соответствующая подматрица  $A^* \subset A$ ;
- на основе подматрицы  $A^*$ , а в дискриминантном анализе - подматриц, однозначно традиционными методами определяются коэффициенты следующей модели прогноза

$$Y^* = \sum_{e=1}^p \beta_e \cdot g(X, \alpha_{ve}) + \beta_0 ; \quad (10)$$

- по построении модели (10) осуществляется прогноз в точке  $x$ .

Заметим, что те подмодели  $g(X, \alpha_{jq})$ , ковариация которых с выбранными компонентами  $\{g(X, \alpha_{ve})\}$  ранее не была достоверно определена, включаются в модель (10) независимо с весом, пропорциональным их величине  $\delta_{jq}$ . Таким образом, в данном случае мы не имеем какой-то одной, заранее рассчитанной модели прогноза. Здесь коллектив из подмоделей  $\{g(X, \alpha_{ve})\}$  формируется и рассчитывается каждый раз при прогнозе.

Использование коллектива кусочно-простых моделей типа (10) увеличивает точность прогноза по сравнению с использованием одной модели и в значительной степени "ослабляет эффект разрывности" функции прогноза типа (I) на границах областей.

## Л и т е р а т у р а

1. КОТЮКОВ В.И. Оптимизация кусочного представления сложных моделей процессов и явлений. - В кн.: Методы представления и аппаратный анализ случайных процессов и полей. Ч.1.-Л., 1974, с.118-122.

2. ЗАГОРУЙКО Н.Г. Методы распознавания и их применение. -М.: Сов.радио, 1972.

3. ВАЛНИК В.Н., МИХАЛЬСКИЙ А.И. О поиске зависимостей методом упорядоченной минимизации риска. -Автоматика и телемеханика, 1974, №10, с.86-96.

4. КОРБУТ А.А., ФИНКЕЛЬШТЕЙН Д.Д. Дискретное программирование. -М.: Наука, 1969.
5. КОТЮКОВ В.И., ТУМИЛОВИЧ В.М. Иерархические модели прогнозирования в АСУТП. -В кн.: Применение математических методов в управлении производственными процессами.- Новосибирск, 1978, с.40-46.
6. ЛБОВ Г.С. Логические функции в задачах эмпирического предсказания. -В кн.: Эмпирическое предсказание и распознавание образов. (Вычислительные системы, вып. 76.) Новосибирск, 1978, с.34-64.
7. КОТЮКОВ В.И. Синтез производственных функций при комплексных статистических исследованиях. -В кн.: Применение математических методов в управлении производственными процессами. - Новосибирск, 1978, с.47-53.
8. РАСТРИГИН Л.А., ЭРЕНШТЕЙН Р.Х. Коллектив алгоритмов для обобщения алгоритмов решения задач. -Техническая кибернетика, 1978, №2, с.116-126.

Поступила в ред.-изд.отд.  
20 ноября 1979 года