

# **Согласованный алгоритм упорядоченной классификации**

**Чижов С.А.**

*(Институт проблем управления РАН; Москва)*

## **1. Введение**

Использование экспертов для получения оценок сложных объектов социально-экономической природы обычно предполагает построение модели объекта оценки, включающей в себя набор исходных критериев, влияющих на состояние объекта, а также алгоритм агрегирования исходных критериев в итоговый показатель. Однако, во многих случаях в задачах такого типа требуется разбить объекты оценки на несколько групп, объекты в каждой из которых близки по некоторому обобщенному показателю, и это разбиение оказывается достаточным для применения к объекту соответствующих управляющих воздействий. Также существенным для небольшого числа групп является то обстоятельство, что арсенал средств управления, находящийся в распоряжении руководителя, ограничен и имеет, как правило, дискретный характер.

Для исследования крупномасштабной системы, состоящей из большого количества объектов и описываемой большим числом характеристик, связанного с необходимостью разбиения системы на подсистемы и изучения характеристик этих подсистем, предназначены методы классификационного анализа данных [1]. Сущность этих методов заключается в нахождении разбиения множества объектов системы на группы схожих. Одним из наиболее разработанных в этой области является вариационный подход [2,3], заключающийся во введении критерия качества классификации, экстремизация которого и приводит к искомому разбиению исходного множества объектов.

В то же время существующие методы классификационного анализа приводят к получению "неупорядоченного" набора классов, тогда как для управления социально-экономическими системами нередко требуется упорядоченность классов объектов в соответствии с некоторым качественным признаком (например, по принципу "лучше-хуже" и т.п.).

В данной работе предлагается постановка задачи классификации с упорядочением классов и рассматривается конкретный алгоритм решения этой задачи.

## 2. Задача упорядоченной классификации

Рассмотрим "стандартную" формальную постановку задачи в рамках вариационного подхода [3].

Пусть задано некоторое множество  $X$  с вероятностной мерой  $P(A)$  ( $A \subseteq X$ ). Размытая классификация  $H$  множества  $X$  на  $g$  классов задается вектор-функцией  $H(X) = (h_1(x), h_2(x), \dots, h_g(x))$ , где  $h_j(x)$  - функция принадлежности точки  $x$   $j$ -му классу ( $g$  - заранее заданное число классов). При этом считается, что  $h_j(x) \in L_2(X, P)$  и значение  $H(X)$  для любой точки  $x$  принадлежит ограниченному замкнутому множеству  $V$  пространства значений функции  $H(x)$ . Множество  $V$  задает тип размытой классификации, рассматриваемой в конкретной задаче. Класс допустимых классификаций обозначается через  $E(V)$ . За критерий качества классификации принимается некоторый функционал  $\Phi(H)$ .

Задача построения размытой классификации состоит в максимизации  $\Phi(H)$  на множестве  $E(V)$ .

Наиболее популярным в прикладных исследованиях является простейший вариант этого функционала - функционал средневзвешенной дисперсии:

$$J_1 = - \sum_{j=1}^g \int_X (x - u_j)^2 h_j(x) dP(x), \quad (1)$$

где  $u_j$  - центр  $j$ -го класса, определяемый по формуле

$$u_j = \frac{\int_X x h_j(x) dP(x)}{\int_X h_j(x) dP(x)}. \quad (2)$$

При подобной постановке задачи все классы предполагаются равнозначными с точки зрения критерия. В то же время в рассматриваемом множестве объектов могут присутствовать объекты достаточно удаленные от всех классов. Поскольку отнесение таких объектов к основным классам носит случайный характер и способно исказить

классификацию остальных объектов, целесообразно использовать так называемый фоновый класс для группирования удаленных объектов. В работе [4] рассмотрена постановка задачи размытой классификации с фоновым классом для вариационного подхода. В этом случае функционал качества классификации выглядит следующим образом:

$$J_2 = - \sum_{j=1}^r \int_X (x - u_j)^2 h_j(x) dP(x) - B \int_X h_0(x) dP(x), \quad (3)$$

где величина  $B$  является некоторой положительной константой, с помощью которой можно варьировать степень влияния фонового класса на оптимальную классификацию. Сам фоновый класс введен с нулевым номером, следовательно вектор-функция принадлежности имеет вид  $H(X) = (h_0(x), h_1(x), \dots, h_r(x))$ , где  $h_0(x)$  - степень принадлежности фоновому классу.

Под упорядоченной классификацией будем понимать разбиение исходных данных на классы, которые определяются центрами, одинаково упорядоченными относительно любой из исходных переменных, т.е.  $u_{jk} < u_{lk}$ ,  $1 \leq j \leq r$ ;  $1 \leq k \leq m$ . Таким образом, можно говорить, что на множестве классов разбиения задан порядок. Требование к самой упорядоченной классификации сводится к следующему: объект, принадлежащий классу  $j$ , должен превосходить по каждой из переменных объекты из классов  $j'$ :  $1 \leq j' < j$ , и, напротив, уступать объектам, принадлежащим классам  $j''$ :  $j < j'' \leq r$ .

Именно классификацию с фоновым классом можно использовать для выявления из заданного множества тех объектов, которые не удовлетворяют предположению об упорядоченности. Однако второе слагаемое в выражении (3) "не реагирует" на неупорядоченность объекта: коэффициенты при весах принадлежности фоновому классу одинаковы для всех объектов. Таким образом, требуется коррекция функционала  $J_2$ .

### 3. Согласованный алгоритм упорядоченной классификации

Для выявления неупорядоченных объектов в данной работе предлагается рассматривать одномерные классификации того же множества объектов. Можно предположить, что объекты, в большей степени отвечающие предположению об упорядоченности, должны иметь более близкие (согласованные друг с другом) одномерные клас-

сификации. Следовательно искомая коррекция функционала  $J_2$  сводится к введению функциональной зависимости от одномерных классификаций для каждого объекта и функционал приобретает вид:

$$J_3 = - \sum_{j=1}^r \int_X (x - u_j)^2 h_j(x) dP(x) + B \int_X f(h^{(1)}, h^{(2)}, \dots, h^{(m)}) h_0(x) dP(x), \quad (4)$$

где  $h^{(k)}$  – вектор-функция принадлежности при одномерной классификации на основе  $k$ -го критерия.

В качестве функции  $f$ , выступающей в роли критерия упорядоченности, можно предложить следующие:

$$a) f(h^{(1)}, h^{(2)}, \dots, h^{(m)}) = \sum_{k=1}^m \sum_{l=1, l \neq k}^r \left[ 2 - \sum_{j=1}^r |h_j^{(k)}(x) - h_j^{(l)}(x)|^t \right], \quad (5)$$

$$b) f(h^{(1)}, h^{(2)}, \dots, h^{(m)}) = 2 - \sum_{j=1}^r \left| \frac{1}{m} \sum_{k=1}^m h_j^{(k)}(x) - h_j(x) \right|^t, \quad (6)$$

где  $t \geq 1$ .

Далее приводится алгоритм для экстремизации данного функционала в случае функции упорядоченности (5). При этом, используется следующий вариант множества  $V$ :

$$V : h_j(x) \geq 0; \sum_{j=0}^r (h_j(x))^2 = 1,$$

который приводит, как показано в работе [3], к размытой классификации.

1) Выбрать произвольные исходные центры кластеров  $u_1, u_2, \dots, u_r$ , которые удовлетворяют предположению об упорядоченности:  $u_{jk} < u_{lk}$ ,  $1 \leq j < l \leq r$ ;  $1 \leq k \leq m$ .

2) Рассчитать по следующим формулам весовые коэффициенты принадлежности объектов для одномерных размытых классификаций без фонового класса на основе соответствующих компонент вектора  $x$ :

$$h_i^{(k)}(x) = 1, h_j^{(k)}(x) = 0, j \neq s, \text{ если } x_k = u_{sk}, 1 \leq k \leq m;$$

$$h_i^{(k)}(x) = 1, h_j^{(k)}(x) = 0, 2 \leq j \leq r, \text{ если } x_k < u_{lk};$$

$$h_i^{(k)}(x) = 1, h_j^{(k)}(x) = 0, 1 \leq j \leq r-1, \text{ если } x_k > u_{rk};$$

$$h_{j-1}^{(k)}(x) = \frac{\left[ (x_k - u_{j-1,k})^2 \right]^{\frac{1}{\lambda-1}}}{\left\{ \left[ (x_k - u_{j-1,k})^2 \right]^{\frac{1}{\lambda-1}} + \left[ (x_k - u_{jk})^2 \right]^{\frac{1}{\lambda-1}} \right\}^{\frac{1}{\lambda}}}$$

$$h_j^{(k)}(x) = \frac{\left[ (x_k - u_{jk})^2 \right]^{\frac{1}{\lambda-1}}}{\left\{ \left[ (x_k - u_{j-1,k})^2 \right]^{\frac{1}{\lambda-1}} + \left[ (x_k - u_{jk})^2 \right]^{\frac{1}{\lambda-1}} \right\}^{\frac{1}{\lambda}}},$$

$$h_s^{(k)}(x) = 0, s \neq j-1, j, \text{ если } u_{j-1,k} < x_k < u_{jk};$$

для всех  $k: 1 \leq k \leq m$ .

3) Найти весовые коэффициенты объектов для основной классификации с фоновым классом:

$$h_s(x) = 1, h_j(x) = 0, j \neq s, \text{ если } x_k = u_{sk}, 1 \leq k \leq m; \quad (7.1)$$

иначе

$$h_s(x) = \frac{\left[ Bf(x) \right]^{\frac{1}{\lambda-1}}}{\left\{ \sum_{j=1}^r \left[ (x - u_j)^2 \right]^{\frac{1}{\lambda-1}} + \left[ Bf(x) \right]^{\frac{1}{\lambda-1}} \right\}^{\frac{1}{\lambda}}}; \quad (7.2)$$

$$h_j(x) = \frac{\left[ (x - u_j)^2 \right]^{\frac{1}{\lambda-1}}}{\left\{ \sum_{j=1}^r \left[ (x - u_j)^2 \right]^{\frac{1}{\lambda-1}} + \left[ Bf(x) \right]^{\frac{1}{\lambda-1}} \right\}^{\frac{1}{\lambda}}}, 1 \leq j \leq r; \quad (7.3)$$

для всех  $i: 1 \leq i \leq n$ .

4) Найти новые координаты для центров кластеров в соответствии с формулой (2).

5) Повторить шаги (2) и (3).

6) Сравнить новую и предыдущую матрицы весовых коэффициентов. Если принятное условие останова выполнено, алгоритм прекращает свою работу. В противном случае требуется повторить шаги (4-6).

Использование другого множества  $V$  требует лишь изменения формул (7.1-7.3). В частности, можно использовать ограничения для вектор-функций принадлежности

$H(x)$ , приводящие к четким классификациям или к классификациям с размытыми границами между классами. Работа данного алгоритма, в частности, его сходимость, была проверена на ряде наборов модельных данных.

Аналогичный алгоритм можно предложить для функционала, в котором использован критерий упорядоченности вида (6).

#### 4. Заключение

В данной работе рассмотрена задача разбиения исходного множества объектов на упорядоченные классы в рамках вариационного подхода к классификационному анализу данных. Для получения упорядоченной классификации используется связь между “упорядоченностью” объекта и согласованностью векторов его принадлежности для одномерных классификаций заданного множества. В работе предложен алгоритм экстремизации для соответствующего функционала.

Как указывалось во введении, данная задача возникает при построении агрегированной оценки на основе экспертных суждений для разбиения множества объектов на упорядоченные классы. Другим важным приложением является выявление групп критериев, на которых исходное множество объектов обладает свойством упорядоченности, что фактически соответствует выделению существенных факторов на множестве исходных показателей.

#### Литература:

- [1] Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. М.: Наука, 1983.
- [2] Bejdec J.C. Pattern recognition with fuzzy objective function algorithms. N.Y., L.: Plenum Press, 1980.
- [3] Бауман Е.В. Методы размытой классификации (вариационный подход) // Автоматика и телемеханика. 1988. № 12. С. 143-156.
- [4] Бауман Е.В., Кузин М.Г. Метод построения размытой классификации с фоновым классом //Методы сбора и анализа сложноорганизованных данных. М.: Институт проблем управления РАН. 1991. С. 95-100.