УДК 621.391

ВВЕДЕНИЕ И ОПТИМИЗАЦИЯ ПАРАМЕТРОВ ИЕРАРХИЧЕСКОЙ ФОРМЫ ПРЕДСТАВЛЕНИЯ ФУНКЦИИ ПРАВДОПОДОБИЯ ПРИ КЛАССИФИКАЦИИ ЛОКАЦИОННОЙ ИНФОРМАЦИИ

к.т.н. В.Е. Саваневич, Е.В. Ветлугин
(представил д.т.н., проф. Ю.И. Лосев)

Вводится иерархическая форма записи функции правдоподобия для относительного описания выборки и предлагается процедура оптимизации её параметров.

При прогрессирующем росте сложности целевой обстановки и ухудшения условий её оценки повышается актуальность задачи создания статистических процедур классификации, обладающих наименьшей сложностью [1] при обеспечении заданного качества принимаемых решений.

В общем случае задача классификации наблюдений формулируется следующим образом. В зоне действия информационной системы находится **Q** объектов. Известны их координаты $\mathbf{X=\{x_1, ..., x_Q\}}$. Получено измерение координат одного из **Q** объектов **y**, которое с априорными вероятностями $\mathbf{p(x_j)}$ принадлежит соответственно каждому из них. Ошибки измерений координат объектов подчинены нормальному распределению с нулевыми средними и дисперсиями $\mathbf{D=\{D_1, ..., D_Q\}}$. Разработчику необходимо синтезировать правило отнесения полученного измерения к одному из **Q** объектов. Желательно, чтобы данное правило обладало, по возможности, меньшей сложностью [1]. Данное требование имеет удобное научно - методическое основание при введении иерархической формы представления функции правдоподобия выборки. В свою очередь, иерархическая форма функции правдоподобия (ФП) бесконечно многообразна как по числу уровней иерархии, так и по числу и местоположению областей разбиений пространства наблюдений $\mathbf{\Omega}$. В связи с этим актуальна задача выбора на каждом шаге иерархии данных параметров.

В дальнейшем в статье рассматриваются только разбиения областей пространства наблюдений (ПН) на две непересекающиеся подобласти (бинарные разбиения) на каждом уровне иерархии.

В соответствии с общим подходом к синтезу алгоритмов минимальной сложности [1] в качестве критерия оптимизации используется критерий максимума средней взаимной информации между пространствами состояний и наблюдений на каждом уровне иерархии [2]:

$$\mathbf{I_{XY} = \sum_{j=1}^{Q} \sum_{i=0}^{1} p(x_j) P(\tilde{y}_i / x_j) \log \frac{P(\tilde{y}_i / x_j)}{P(\tilde{y}_i)}}, \qquad (1)$$

© к.т.н. В.Е. Саваневич, Е.В. Ветлугин, 2002

где $\tilde{y}_0, \tilde{y}_1$ – события, которые состоят в попадании измерения y в области ПН D_1 и D_0 $(D_1 \bigcup D_0 = \Omega, D_1 \bigcap D_0 = \varnothing$ $\tilde{y}_0 + \tilde{y}_1 = 1,\ \tilde{y}_1 = 1 - \tilde{y}_0)$; $P(\tilde{y}_i) = \sum_{x=1}^{Q} p(x_j) P(y_i / x_j)$ – безусловная вероятность попадания измерения в D_i - ую область $(p(\tilde{y}_0) + p(\tilde{y}_1) = 1)$; $P(\tilde{y}_i / x_j) = \int_{y \in D_i} f(y / x_j) dy$ – вероятность попадания измерения от j-го объекта в D_i - ую область ПН $(P(\tilde{y}_0 / x_j) + P(\tilde{y}_1 / x_j) = 1)$; $f(y / x_j)$ – плотность распределения измерения координат j - го объекта.

Использование информации между состояниями x_j и наблюдениями $\tilde{y}$ связано, прежде всего, с тем, что разработчику необходимо не абсолютное описание выборки, а относительное – т.е. описание классифицирующих признаков, составляющих выборку или её часть [3].

Итак, обычно закон распределения случайных величин (СВ) представляется в виде совокупности вероятностей событий (дискретные СВ) или в виде аналитических выражений для плотностей распределений (непрерывные СВ). При дискретизации непрерывных СВ значения вероятностей событий (вероятностей попадания СВ в заданные интервалы) в однокоординатном случае определяются интегралами вида

$$P(\tilde{y}_i / x_j) = \int_{\alpha_i}^{\beta_i} f(y / x_j) dy\,, \qquad (2)$$

где α_i, β_i - координаты начала и конца i - го интервала.

Вместе с тем при синтезе иерархических решающих правил (РП) удобна иерархическая форма представления закона распределения СВ. Сущность такой формы представления заключается в следующем (рис.1). Вероятность попадания СВ y, например, в третий интервал P_3 может быть представлена в виде

$$P_3 = P(\tilde{y}_3 / x_j) = P(\tilde{y}_{3,4} / x_j) \nu(\tilde{y}_3 / x_j, \tilde{y}_{3,4}) = P_{3,4}\, \nu_3\,,$$

где $P_{n,m} = P(\tilde{y}_{n,m} / x_j)$ - вероятность попадания $(\tilde{y}_{n,m} = 1)$ измерения y от объекта j в интервал с номерами дискрет $\lambda = \overline{n, m}$; $\nu_{n,\kappa} = \nu(\tilde{y}_{n,\kappa} / x_j, \tilde{y}_{n,m})$ - вероятность попадания $(\tilde{y}_{n,\kappa} = 1)$ измерения y от j-го объекта в интервал с номерами дискрет $\lambda_\kappa = \overline{n, \kappa}$, при условии, что оно попадает в интервал с номерами дискрет $\lambda_n = \overline{n, m}$. Вероятность $P_{3,4} = P(\tilde{y}_{3,4} / x_i)$ равна

$$P_{3,4} = P(\tilde{y}_{3,4} / x_j) = P(\tilde{y}_{1,4} / x_j) \nu(\tilde{y}_{3,4} / x_j, \tilde{y}_{1,4}) = P_{1,4}\, \nu_{3,4}\,.$$

Наконец, если ввести ещё один уровень мультипликативной иерархии, то вероятность $\mathbf{P(\tilde{y}_{1,4} / x_j)}$ примет вид

$$\mathbf{P_{1,4} = P(\tilde{y}_{1,8} / x_j)\,\nu(\tilde{y}_{1,4} / x_j, \tilde{y}_{1,8}) = P_{1,8}\,\nu_{1,4}}.$$

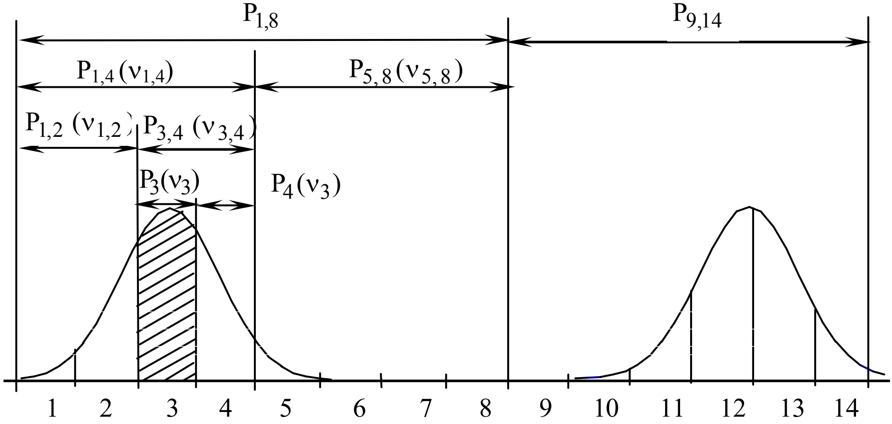

Рис. 1. К пояснению понятия иерархической формы представления ФП

Запишем общее выражение для вероятности попадания СВ **y** в третий интервал

$$\mathbf{P_3 = P(\tilde{y}_3 / x_j) = P(\tilde{y}_{1,8} / x_j)\,\nu(\tilde{y}_{1,4} / x_j, \tilde{y}_{1,8})\,\nu(\tilde{y}_{3,4} / x_j, \tilde{y}_{1,4})\,\nu(\tilde{y}_3 / x_j, \tilde{y}_{3,4}) =}$$
$$\mathbf{= P_{1,8}\,\nu_{1,4}\,\nu_{3,4}\,\nu_3}.$$

Последнее выражение в развёрнутом виде с учётом (2) можно переписать следующим образом:

$$\mathbf{P_3 = P(\tilde{y}_3 / x_j) =}$$

$$\mathbf{= \int_{\alpha_1}^{\beta_8} f(y / x_j)dy \frac{\int_{\alpha_1}^{\beta_4} f(y / x_j)dy}{\int_{\alpha_1}^{\beta_8} f(y / x_j)dy} * \frac{\int_{\alpha_3}^{\beta_4} f(y / x_j)dy}{\int_{\alpha_1}^{\beta_4} f(y / x_j)dy} * \frac{\int_{\alpha_3}^{\beta_3} f(y / x_j)dy}{\int_{\alpha_3}^{\beta_4} f(y / x_j)dy} = \int_{\alpha_3}^{\beta_3} f(y / x_j)dy}.$$

В общем случае СВ **Y** заменяется совокупностью СВ $\boldsymbol{\zeta_1, \ldots, \zeta_{n(y)}}$, где $\boldsymbol{\zeta_\kappa}$ - номер интервала, в который попала СВ на **к**-м уровне иерархии. Число $\mathbf{n = n(y)}$ СВ типа ζ, которыми заменяется СВ **Y**, зависит от конкретного значения **y** (аналог неравномерного кодирования). В описанном примере

$$\mathbf{n(y \in [\alpha_3, \beta_3]) = 4;\ \zeta_1 = \tilde{y}_{1,8};\ \zeta_2 = \tilde{y}_{1,4};\ \zeta_3 = \tilde{y}_{3,4};\ \zeta_4 = \tilde{y}_3}.$$

Случайные величины $\boldsymbol{\zeta_\kappa\ (\kappa \neq n)}$ являются дискретными (в частном случае – бинарными). Случайная величина $\boldsymbol{\zeta_n}$ может быть как дискретной (для описания вероятностей попадания в интервал), так и непрерывной (для представления плотности распределения).

В соответствии со свойством иерархической мультипликативности условных вероятностей [4] вероятность попадания СВ **Y** в любой интервал представляется выражением:

$$P_i = P_{1i} \prod_{\kappa=2}^{n(i)} \frac{P_{\kappa i}}{P_{(\kappa-1)i}} = P_{1i} \prod_{\kappa=2}^{n(i)} \nu_{\kappa i}\,, \tag{3}$$

где P_{1i}, $\nu_{\kappa i}$ - безусловная и соответствующие условные вероятности попадания СВ **Y** в интервалы, включающие в себя **i** - й интервал на **к** - м уровне иерархии.

Иерархическая форма представления функции правдоподобия достаточно наглядно демонстрирует различие между абсолютным и относительным описанием выборки. Так в рамках информационного подхода [1] сложность абсолютного описания эквивалентна сложности пространства наблюдений, т.е. равна его энтропии H_Y. В свою очередь для различения простой гипотезы против простой альтернативы в примере, приведенном на рис. 1 достаточно индикатора попадания в интервал $i = \overline{1,4}$ или $i = \overline{5,8}$. Энтропия последнего (минимально необходимое относительное описание) никогда не превысит единицы. Следовательно, энтропия абсолютного описания выборки превосходит энтропию относительного на величину энтропии нормального закона $H_0 = 0.5\log(2\pi e D)$ или при σ = **1, 5, 10** в 3.047, 5.369, 6.369 раз соответственно.

Понятие иерархической формы представления введено. Далее следует выбрать параметры гиперповерхности, разделяющей пространство или подпространство наблюдений на область D_1 и дополнение до неё на каждом уровне иерархии. Для обеспечения максимума выбранного критерия (1) последние должны на каждом уровне удовлетворять условию:

$$\begin{aligned}\frac{\partial I_{XY}}{\partial D_1} = \sum_{j=1}^{Q} \sum_{i=0}^{1} p(x_j) \Bigg[\frac{\partial P(\tilde{y}_i / x_j)}{\partial D_1} \left(1 + \log \frac{P(\tilde{y}_i / x_j)}{P(\tilde{y}_i)} \right) - \\ - \frac{P(\tilde{y}_i / x_j)}{P(\tilde{y}_i)} \sum_{j=1}^{M} p(x_j) \frac{\partial P(\tilde{y}_i / x_j)}{\partial D_1} \Bigg] = 0.\end{aligned} \tag{4}$$

В однокоординатном случае разделяющая гиперповерхность вырождается в порог **z**, а выражения для производных, входящих в (4), имеют вид:

$$\frac{dP(\tilde{y}_0 / x_j)}{dz} = \frac{dP(0 / x_j)}{dz} = \frac{\partial}{\partial z} \int_{-\infty}^{z} f(y / x_j) dy = f(z / x_j);$$

$$\frac{dP(\tilde{y}_1 / x_j)}{dz} = \frac{dP(1 / x_j)}{dz} = \frac{\partial}{\partial z} \int_{z}^{-\infty} f(y / x_j) dy = -f(z / x_j).$$

При этом условие (4) также упростится:

$$\frac{dI_{XY}}{dz} = \tag{5}$$

$$= \sum_{j=1}^{Q} p(x_j)\left[f(z/x_j)\left(\log\frac{P(\tilde{y}_1)}{P(\tilde{y}_0)} - \log\frac{P(\tilde{y}_1/x_j)}{P(\tilde{y}_0/x_j)} \right) + G(z)\left(\frac{P(\tilde{y}_1/x_j)}{P(\tilde{y}_1)} - \frac{P(\tilde{y}_0/x_j)}{P(\tilde{y}_0)} \right)\right] = 0,$$

где $G(z) = \sum_{j=1}^{Q} p(x_j) f(z/x_j)$.

Выражение (5) является необходимым, но не достаточным условием экстремального бинарного разбиения. Графики безусловной плотности вероятности выборки $G(z)$, энтропии индикатора $H_{\tilde{Y}}$, взаимной информации $I_{XY}(z)$ и её производной $dI_{XY}(z)/dz$ при $x_1 = 4.0$, $x_2 = 14.0$, $x_3 = 18.0$, $x_4 = 22.0$, $x_5 = 32.0$, $\sigma_j = \sigma = 1$ для $j = \overline{1,5}$, а также энтропии индикатора $H_{\tilde{Y}} = -p(0)\log p(0) - p(1)\log p(1)$ приведены на рис. 2.

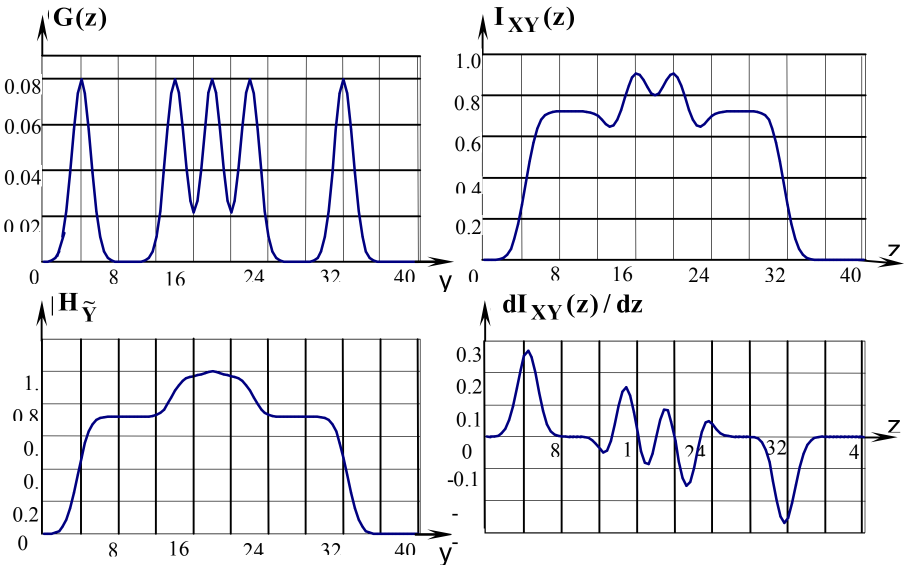

Рис. 2. Графики функций $G(z), H_{\tilde{Y}}, I_{XY}(z), dI_{XY}(z)/dz$

Видно, что функции $I_{XY}(z)$ и $dI_{XY}(z)/dz$ имеют ряд особенностей, которые не позволяют использовать простейшие вычислительные схемы для нахождения глобального максимума $I_{XY}(z)$.

Проведенные исследования показали, что максимум взаимной информации расположен в точках, которые являются границами максимальности дискриминантных информантов [5] объектов классификации при простой функции потерь или в их окрестностях. Координаты данных границ можно определить из условия

$$p(x_n)\, f(z/x_n) = p(x_m)\, f(z/x_m)\,. \tag{6}$$

При конкретизации на случай нормальных распределений со средними m_j, m_i и дисперсиями D_j, D_i уравнение (6) можно записать в виде

$$\frac{p(x_n)}{p(x_m)}\sqrt{\frac{D_m}{D_n}}\exp\left(-\frac{(z-x_n)^2}{2\sigma_n^{\,2}}+\frac{(z-x_m)^2}{2\sigma_m^{\,2}}\right)=1\,. \tag{7}$$

Корни (7) определяются как $z_{1,2} = (-b \pm \sqrt{b^2 - 4ac})/2a$, где $a = D_n - D_m$;

$b = 2(x_n D_m - x_m D_n)$; $c = x_m^2 D_n - x_n D_m + 2 D_m D_n \ln \dfrac{p(x_n)}{p(x_m)}\sqrt{\dfrac{D_m}{D_n}}$.

В общем случае всего будет иметь место $Q(Q-1)/2$ уравнений типа (7) и только незначительная часть их корней (в случае примерно одинаковых дисперсий – $Q-1$) является границами максимальности дискриминантных информантов.

Для отбора границ из множества корней целесообразно воспользоваться следующим селективным признаком. Корень уравнения (7) является границей максимальности дискриминантного информанта j-го объекта, если $\forall\, \kappa = \overline{1,Q}$, $\kappa \neq j$ выполняется $p(x_j) f(z/x_j) \geq p(x_\kappa) f(z/x_\kappa)$.

В дальнейшем необходимо выбрать ту границу, которая имеет максимальное значение взаимной информации между состояниями и наблюдениями. Расчёт порогов разбиений $z_{i\kappa}$ для i-х подинтервалов на последующих шагах осуществляется аналогично. Следует обратить внимание на область применимости предлагаемой вычислительной схемы. Во-первых, такая схема применима на первых **Q-1** шагах при примерно одинаковых дисперсиях и на несколько большем числе шагов при наличии компонент с существенно отличающимися дисперсиями.

Число правильных шагов при этом может сократиться на один при наличии объекта с априорной вероятностью, в сотни раз меньшей априорных вероятностей других объектов, при достаточной близости других объектов друг от друга. Хотя сложно найти практически значимый случай, который соответствует такой ситуации.

В дальнейшем (на Q - м и последующих шагах) интервалы анализа сокращаются, функция $I_{XY}(z)$ на них становится унимодальной, а её производная - почти линейно переходит через ноль в точках экстремума. Тем самым проблема поиска экстремальных разбиений подинтервалов превращается в простейшую стандартную процедуру. Кроме того, для относительного описания выборки необходимость в таких дальнейших разбиениях может просто отсутствовать.

При существенном отличии дисперсий разных компонент аргумент максимума информации уже не находится в одной из точек, соответствующих границам максимальности дискриминантных информантов. Вместе с тем аргумент максимума достаточно близок к ним и их можно использовать в качестве хороших начальных приближений. Расстояние между таким начальным приближением и аргументом максимума средней взаимной информации тем больше, чем ближе объекты расположены один к другому. Особенно при равных априорных вероятностях появления измерений от объектов. Максимальное значение этого расстояния при $\mathbf{(m_{i+1} - m_i)/\sqrt{D_i} = 1}$ **и** $\mathbf{D_{i+1}/D_i = 4}$ не превосходит $\mathbf{0.6\sqrt{D_i}}$. При этом для поиска экстремумов функции $\mathbf{I_{XY}}$ при использовании в качестве начальных приближений границ максимальности дискриминантных информантов достаточно использовать метод Ньютона.

Итак, в работе введена иерархическая форма записи функции правдоподобия в целях относительного описания выборки. Дополнительно предложена вычислительная процедура, позволяющая найти параметры данной формы, оптимальные, по критерию максимума средней взаимной информации между состояниями и решениями на каждом уровне иерархии.

ЛИТЕРАТУРА

1. *Саваневич В.Е. Информационный подход к синтезу статистических алгоритмов с минимальной сложностью // Системи обробки інформації. – Харків : НАНУ, ПАНМ, ХВУ. – 2000. – Вип. 3(9). – С. 123 - 128.*
2. *Колесник В.Д., Полтырев Г.Ш. Курс теории информации. – М.: Наука, 1982. – 416 с.*
3. *Барабаш Ю.Л., Варский Б.В., Зиновьев В.Т. Вопросы статистической теории распознавания. – М.: Сов. радио, 1967. – 400 с.*
4. *Стратонович Р.Л. Теория информации. – М.: Сов. радио, 1975. – 424 с.*
5. *Рао С.Р. Линейные статистические методы и их применения. – М.: Наука, 1968. – 548 с.*

Поступила 24.12.2001

***САВАНЕВИЧ Вадим Евгениевич**, канд. тех. наук, доцент, докторант Харьковского военного университета. В 1986 году закончил Житомирское высшее училище радиоэлектроники ПВО. Область научных интересов – алгоритмы минимальной сложности, применение теории информации в локационных системах, информметрия.*

***ВЕТЛУГИН Евгений Владимирович**, адъюнкт Харьковского военного университета. В 1997 году закончил Харьковский военный университет. Область научных интересов – иерархические классификаторы локационной информации, программирование.*