

К.Т. Протасов

ПАРАМЕТРИЗАЦИЯ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ ДЛЯ РАСПОЗНАВАНИЯ ОБРАЗОВ, ОСНОВАННАЯ НА НОРМАЛИЗУЮЩИХ ПРЕОБРАЗОВАНИЯХ

В рамках байесова подхода к построению решающих правил распознавания образов в пространстве описаний высокой размерности решается задача восстановления многомерных условных функций плотности классов на основе нормализующих преобразований, удовлетворяющих двум условиям. Первое из них состоит в том, чтобы оценки распределений отдельных признаков с достаточной для практики надежностью согласовывались с истинными одномерными распределениями. Второе требование сводится к тому, чтобы аппроксимирующее распределение описывало статистические связи между компонентами вектора наблюдений.

При тематическом дешифрировании аэрокосмической видеоинформации одной из центральных является задача сегментации спектро-зональных изображений подстилающей поверхности и облачности. При этом оцифрованное исходное изображение декомпозируется на отдельные фрагменты, которые являются в данном случае наблюдаемыми сигналами. Алгоритм распознавания образов, выраженный в виде решающего правила, по наблюдаемому сигналу идентифицирует ситуацию, породившую соответствующий класс наблюдаемых данных.

Среди многочисленных подходов к построению решающих правил распознавания образов наиболее формализованным является байесов подход, связанный с проверкой статистических гипотез. Именно в рамках этого подхода, имеющего в качестве методологической основы теорию проверки статистических гипотез и теорию синтеза адаптивных информационных систем, в условиях полной априорной информации получены оптимальные решающие правила принятия решений из условий минимума функции потерь. Однако непосредственное использование этого подхода сопряжено со значительными трудностями, и в основном ввиду того, что для большинства реальных задач распознавания образов априорная информация неполна, так как нам неизвестны ни априорное распределение образов (классов), ни, что более существенно, условные функции плотности, описывающие распознаваемые ситуации.

Критерий качества алгоритма распознавания определяется усреднением платежной матрицы. Элементы этой матрицы задают достаточно произвольно, лишь обеспечивая свойство выпуклости функции потерь. Поэтому вполне допустимым является задание априорных вероятностей из условия максимума априорной неопределенности, руководствуясь постулатом Лапласа–Байеса. Условные же функции плотности распределения вероятностей необходимо восстанавливать, имея в распоряжении лишь обучающие выборки классов, как правило, скудного объема. Кроме того, в задачах, связанных с анализом изображений, наблюдаемые сигналы имеют высокую размерность, превосходящую объем выборочных данных.

В связи с этим становится актуальным рассмотреть задачу восстановления многомерных условных функций плотности по выборочным данным обучающих последовательностей, описывая наблюдаемые величины или их частотные распределения гибкой системой математических формул.

Будем исходить из стандартной постановки задачи распознавания образов в статистической интерпретации [1].

В евклидовом n -мерном пространстве наблюдений E^n , элементами которого являются фрагменты оцифрованных изображений, представленные в виде вектора $\mathbf{X} \in E^n$, и пространстве гипотез $\Lambda = \{1, \dots, L\}$, где L – число гипотез (образов), определены вероятностные меры с априорным распределением классов $P(\lambda)$ и условными функциями плотности вероятностей $f(\mathbf{x}/\lambda)$, $\lambda \in \Lambda$, $\mathbf{x} \in E^n$ и простая матрица потерь от принимаемых решений $(1 - \delta_{\lambda\mu})$, где $\delta_{\lambda\mu}$ – символ Кронекера. Оптимальное в смысле минимума средних потерь байесово решающее правило принятия гипотезы $\lambda \in \Lambda$ из серии взаимоисключающих гипотез имеет вид

$$u(\mathbf{x}) = \arg \max_{\lambda \in \Lambda} P(\lambda) f(\mathbf{x} / \lambda), \quad (1)$$

где решение u также из пространства Λ .

Это решающее правило можно представить в эквивалентной форме в виде отношения правдоподобия, сравниваемого с порогом.

В случае, когда задана выборка наблюдаемых векторов $\mathbf{X}_1^\lambda, \dots, \mathbf{X}_{N_\lambda}^\lambda$, N_λ – объем выборки, $\lambda \in \Lambda$, классифицированных <учителем>, средний риск можно оценить эмпирическим риском следующим образом:

$$R = \sum_{\lambda \in \Lambda} \frac{1}{N_\lambda} \sum_{k=1}^{N_\lambda} P(\lambda) I \{ \lambda = \arg \max_{\mu \in \Lambda} P(\mu) f(\mathbf{X}_k^\lambda / \mu) \}, \quad (2)$$

где $I \{ \text{истина} \} = 0$, $I \{ \text{ложь} \} = 1$ – характеристическая функция; N_λ – объем выборки класса $\lambda \in \Lambda$.

Естественно потребовать, чтобы восстанавливаемые параметрические функции плотности $f(\mathbf{x}/\lambda)$, $\lambda \in \Lambda$ для решающего правила (1) удовлетворяли двум условиям. Первое из них состоит в том, чтобы оценки распределений отдельных признаков с достаточной для практики надежностью согласовывались с истинными одномерными распределениями. Второе требование сводится к тому, чтобы аппроксимирующее распределение в какой-то степени описывало статистические связи между компонентами случайного вектора. В этой связи широкий класс многомерных параметрических распределений можно получить, используя идею преобразования распределения наблюдаемой величины в гауссово распределение.

В общем виде идею использования преобразований случайных величин для конструирования многомерных параметрических функций плотности, обладающих повышенной аппроксимирующей способностью описывать вероятностные свойства выборочных данных, поясним следующим образом.

Необходимо построить определенное с точностью до параметров преобразование

$$\mathbf{y} = \mathbf{y}(\mathbf{x}), \quad (3)$$

которое переводит векторную случайную величину $\mathbf{X} \in E^n$, имеющую распределение $F(\mathbf{x})$ и функцию плотности вероятностей $f(\mathbf{x})$, в случайную векторную величину $\mathbf{Y} \in E^n$, с распределением $G(\mathbf{y})$ и функцией плотности вероятностей $g(\mathbf{y})$. После чего функция плотности случайного вектора \mathbf{X} находится следующим стандартным образом:

$$f(\mathbf{x}) = g(\mathbf{y}(\mathbf{x})) \left| \frac{D \mathbf{y}(\mathbf{x})}{D \mathbf{x}} \right|, \quad (4)$$

где $|D \mathbf{y}(\mathbf{x})/D \mathbf{x}|$ – якобиан преобразования (3).

Аппроксимирующая способность полученной таким образом функции $f(\mathbf{x})$ возрастает, так как она зависит не только от параметров <более простого> семейства $g(\mathbf{y})$, но и от параметров преобразования (3).

Для реализации отображения (3) воспользуемся преобразованиями, осуществляемыми непрерывными интегральными функциями распределения [2, с. 187], которые осуществляют переход от случайного вектора \mathbf{X} к $\mathbf{Y} \in E^n$ с необходимыми свойствами. Рассмотрим упрощенный вариант этого преобразования, представив его в виде следующей двухэтапной процедуры.

На первом этапе определим преобразование, осуществляемое одномерными распределениями, переводящими компоненты случайного вектора \mathbf{X} в соответствующие компоненты вспомогательного вектора \mathbf{Z} , $\mathbf{Z} \in E^n$, причем каждая из компонент вектора \mathbf{Z} распределена равномерно на отрезке [0, 1]

$$Z^i = F_i(X^i), \quad i = 1, \dots, n, \quad (5)$$

где $F_i(X^i)$ – маргинальное распределение компоненты X^i – вектора $\mathbf{X} \in E^n$.

На втором этапе зададим строго возрастающие функции распределений $G_i(y^i)$, такие, что

$$G_i(Y^i) = Z^i, \quad i = 1, \dots, n, \quad (6)$$

где Z^i – вспомогательная случайная величина из выражения (5).

Преобразование (3), обеспечивающее покомпонентный переход от вектора \mathbf{X} к \mathbf{Y} , с учетом (5) и (6) примет вид

$$y^i = G_i^{-1}(F_i(x^i)), \quad i = 1, \dots, n, \quad (7)$$

где $G_i(y^i) = \int_{-\infty}^{y^i} g_i(t) dt$, $i = 1, \dots, n$, $g_i(t)$ – одномерная функция плотности вероятностей компоненты i вектора \mathbf{Y} . Далее, для определенности, функции $G_i(y^i)$, $i = 1, \dots, n$, в формуле (6), (7) будем полагать гауссовыми распределениями.

Чтобы в какой-то степени скомпенсировать несовершенство преобразования (7) и учесть зависимость компонент вектора \mathbf{Y} , совместное распределение компонент Y^1, \dots, Y^n будем считать многомерным гауссовым распределением с матрицей корреляции Σ и вектором средних μ . Тогда искомая функция плотности исходного вектора \mathbf{X} запишется, с учетом (4), (7), в следующем виде:

$$f(\mathbf{x}) = \frac{|\Sigma|^{-1/2}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} (\mathbf{G}^{-1}(\mathbf{F}(\mathbf{x})) - \mu)^T \Sigma^{-1} (\mathbf{G}^{-1}(\mathbf{F}(\mathbf{x})) - \mu) \right\} \prod_{i=1}^n \left\{ \frac{dG_i^{-1}(F_i(x^i))}{dF_i} f_i(x^i) \right\}, \quad (8)$$

где $\mathbf{G}^{-1}(\mathbf{F}(\mathbf{x})) - \mu = (G_i^{-1}(F_i(x^i)) - \mu^i)$, $f_i(x^i) = \frac{dF_i(x^i)}{dx^i}$, $i = 1, \dots, n$; τ – знак транспонирования.

Таким образом, осуществляя покомпонентную нормализацию и учитывая межкомпонентные связи преобразованных данных с помощью корреляций, мы получим параметрическое представление функции плотности вероятностей для описания распознаваемых образов решающим правилом (1).

При использовании выражения (8) следует иметь в виду, что $G_i^{-1}(\cdot)$ – функция, обратная гауссову распределению, которое не имеет простого аналитического вида. Однако известны достаточно точные аппроксимационные выражения как для интеграла вероятностей, так и для обратной к нему функции. Так, λ – распределение, введенное Дж. Тьюки (3), с высокой точностью аппроксимирует обратные функции непрерывных распределений

$$y = G^{-1}(z) = \lambda_1 + (z^{\lambda_3} - (1-z)^{\lambda_3}) / \lambda_2, \quad (9)$$

где $z \in [0, 1]$; среднее и медиана величины Y совпадают и равны λ_1 ; $\lambda_1 = M[Y]$ (M – математическое ожидание); λ_1 – параметр локальности; λ_2 – параметр масштаба; λ_3 – параметр формы.

В частности, для определения обратной функции нормального распределения со средним μ и дисперсией σ^2 имеем

$$\lambda_1 = \mu; \quad \lambda_2 = 0,1975/\sigma; \quad \lambda_3 = 0,1349.$$

Кроме того, для определения многомерной функции плотности $f(\mathbf{x})$ по формуле (8) необходимо восстановить одномерные вероятностные характеристики $F_i(x^i), f_i(x^i)$, $i = 1, \dots, n$.

Для решения этой задачи по выборочным данным естественно воспользоваться системой кривых Пирсона [4] или аппроксимациями выборочных распределений сплайнами.

Описанная процедура построения функций плотности для решающего правила распознавания образов существенно упрощается, если воспользоваться <готовыми> нормализующими преобразованиями [4, 5].

Среди <готовых> нормализующих преобразований следует выделить преобразования, предложенные Джонсоном, позволяющие аппроксимировать широкий класс одномодальных и

бимодальных распределений, встречающихся на практике [5, 6]. Процедуру восстановления параметрической функции плотности в этом случае можно представить в виде двух этапов.

На первом этапе для каждого признака подбирается преобразование Джонсона, с достаточной надежностью согласованное с истинным неизвестным распределением, на втором – для описания статистических связей преобразованных компонент наблюдаемого вектора оцениваются коэффициенты корреляции между признаками. После этого совместное распределение компонент случайного вектора записывается стандартным образом.

Рассмотрим подробнее первый этап. Пусть $X \in E^1$ – случайная величина, для которой требуется подобрать распределение Джонсона. Его можно записать в общем виде:

$$\xi = \gamma + \delta \tau(x; \varepsilon, \lambda), \quad (10)$$

где параметры $\gamma, \delta, \varepsilon, \lambda$ и функция $\tau(\cdot)$ выбираются таким образом, чтобы ξ имела нормальное распределение $N(0, 1)$ с нулевым средним и единичной дисперсией.

Джонсон предложил следующие три формы семейства функций $\tau(\cdot)$:

$$\begin{aligned} S_L : \tau_L(x; \varepsilon, \lambda) &= \ln\left(\frac{x-\varepsilon}{\lambda}\right), \quad x \geq \varepsilon, \\ S_B : \tau_B(x; \varepsilon, \lambda) &= \ln\left(\frac{x-\varepsilon}{\varepsilon+\lambda-x}\right), \quad \varepsilon \leq x \leq \varepsilon + \lambda, \\ S_U : \tau_U(x; \varepsilon, \lambda) &= \ln\left(\frac{x-\varepsilon}{\lambda} + \sqrt{\left(\frac{x-\varepsilon}{\lambda}\right)^2 + 1}\right), \quad -\infty < x < \infty. \end{aligned} \quad (11)$$

Зная эмпирические оценки $\hat{\mu}_i$ истинных центральных моментов μ_i ($i = 2, 3, 4$) исходной случайной величины X , можно заранее определить, какое из семейств функций $\tau(\cdot)$ целесообразней выбрать для описания распределения X , если задана выборка реализаций X_1, \dots, X_N объема N .

Методика выбора функций $\tau(\cdot)$ основана на оценках показателя асимметрии $\beta_1 = \mu_3^2/\mu_2^3$ и относительного показателя эксцесса $\beta_2 = \mu_4/\mu_2^2$ и состоит в следующем. Если в плоскости (β_1, β_2) провести кривую, заданную параметрическим уравнением

$$\begin{cases} \beta_1 = (\omega - 1)(\omega + 2)^2, \\ \beta_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3, \end{cases} \quad (12)$$

то для распределений, у которых значения β_1 и β_2 лежат на линии и вблизи линии, определяемой уравнением (12), следует воспользоваться аппроксимацией S_L (логарифмически нормальным законом); для распределений со значениями β_1, β_2 , лежащими выше линии S_L , – аппроксимацией S_B , и со значениями β_1 и β_2 , лежащими ниже этой линии, – аппроксимацией S_U Джонсона.

Для практического использования (12) удобно выразить ω из первого уравнения следующим образом:

$$\omega^3 + 3\omega^2 - (4 + \beta_1) = 0,$$

и найти вещественный корень

$$\omega_1 = \sqrt[3]{\frac{2 + \beta_1}{2} + \sqrt{\left(\frac{2 + \beta_1}{2}\right)^2 - 1}} + \sqrt[3]{\frac{2 + \beta_1}{2} - \sqrt{\left(\frac{2 + \beta_1}{2}\right)^2 - 1}} - 1.$$

Подставляя ω_1 во второе уравнение (12), определим знак и величину <рассогласования> с β_2

$$\varepsilon_0 = \beta_2 - (\omega_1^4 + 2\omega_1^3 + 3\omega_1^2 - 3). \quad (13)$$

Если теперь в качестве допустимой величины рассогласования возьмем ε_1 , тогда алгоритм выбора семейства распределений можно записать в следующем виде:

если расчетное по формуле (13) $|\varepsilon_0| \leq \varepsilon_1$, то выбираем семейство распределений, связанное с уравнением линии (12), то есть S_L ;

если $|\varepsilon_0| > \varepsilon_1$ и ε_0 – отрицательная величина, то выбираем семейство распределений S_U ;

если $\varepsilon_0 > 0$, то выбираем семейство распределений S_B .

После конкретизации вида функции $\tau(\cdot)$ в выражении (10) распределение Джонсона, описывающее плотность вероятности случайной величины X , будет иметь вид

$$f(x) = \frac{|\delta \tau'_x(x; \varepsilon, \lambda)|}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (\gamma + \delta \tau(x; \varepsilon, \lambda))^2 \right\}, \quad (14)$$

где

$$\tau'_x(x; \varepsilon, \lambda) = d \tau(x; \varepsilon, \lambda) / dx.$$

Предположим, что параметры распределений Джонсона каким-либо образом заданы, на втором этапе процедуры аппроксимации \mathbf{X} – векторная величина, $\mathbf{X} \in \mathbf{E}^n$, $\mathbf{X} = (X^1, \dots, X^n)^T$, каждая компонента которой преобразуется в компоненту вектора $\xi \in E^n$ по формулам (10). Для восстановления межкомпонентных связей случайных величин ξ^i и ξ^j вектора ξ воспользуемся оценками коэффициентов корреляции по материалу обучения $\mathbf{X}_1, \dots, \mathbf{X}_N$; N – объем выборки, $\mathbf{X} \in E^n$:

$$r_{ij} = \frac{1}{N} \sum_{k=1}^N \xi_k^i \xi_k^j, \quad (15)$$

где $x_k^s = \gamma^s + \delta^s \tau(X_k^s; \varepsilon^s, \lambda^s)$; $s = i, j$; $i, j = 1, \dots, n$.

Тогда совместное распределение компонент вектора ξ , в соответствии с предложением о нормальности, запишется следующим образом:

$$f(\xi) = \frac{1}{(2\pi)^{n/2} |R|^{1/2}} \exp \left\{ -\frac{1}{2} \xi^T R^{-1} \xi \right\}, \quad (16)$$

где $R = (r_{ij})$ – матрица корреляции из коэффициентов (15). Отсюда оценка распределения исходного вектора \mathbf{X} имеет вид

$$f(\mathbf{x}) = \frac{\prod_{i=1}^n \delta^i \tau'_x(x^i; \varepsilon^i, \lambda^i)}{(2\pi)^{n/2} |R|^{1/2}} \exp \left\{ -\frac{1}{2} (\gamma + \delta \tau(\mathbf{x}; \varepsilon, \lambda))^T R^{-1} (\gamma + \delta \tau(\mathbf{x}; \varepsilon, \lambda)) \right\}, \quad (17)$$

где вектор

$$\gamma + \delta \tau(\mathbf{x}; \varepsilon, \lambda) = \begin{pmatrix} \gamma^1 + \delta^1 \tau(x^1; \varepsilon^1, \lambda^1) \\ \vdots \\ \gamma^n + \delta^n \tau(x^n; \varepsilon^n, \lambda^n) \end{pmatrix}.$$

Перейдем теперь к вопросам оценивания параметров семейств S_L, S_B распределений Джонсона. Во-первых, заметим, что ε и λ имеют простой смысл: параметр ε – величина нижней границы, а параметр $\varepsilon + \lambda$ – верхней границы значений случайной величины X . Во многих случаях эти параметры можно оценить, исходя из физического смысла измеряемых признаков, а также непосредственно по обучающей выборке. При известных параметрах ε и λ параметры γ и δ определим методом максимального правдоподобия.

Пусть $y = \{X_1, \dots, X_N\}$ независимые выборочные значения случайной величины X , N – объем выборки. Функция правдоподобия с учетом (14) будет иметь следующий вид:

$$L(\gamma, \delta / y) = \left(\frac{\delta}{\sqrt{2\pi}} \right)^N \prod_{j=1}^N \tau'_x(X_j; \varepsilon, \lambda) \exp \left\{ -\frac{1}{2} \sum_{j=1}^N (\gamma + \delta \tau(X_j; \varepsilon, \lambda))^2 \right\}. \quad (18)$$

Дифференцируя $L(\gamma, \delta/y)$ по γ и δ и приравнявая производные к нулю, получим систему уравнений

$$\begin{aligned} \gamma + \delta \frac{1}{N} \sum_{j=1}^N \tau(X_j; \varepsilon, \lambda) &= 0, \\ \delta^2 \frac{1}{N} \sum_{j=1}^N \tau^2(X_j; \varepsilon, \lambda) + \gamma \delta \frac{1}{N} \sum_{j=1}^N \tau(X_j; \varepsilon, \lambda) &= 1, \end{aligned} \quad (19)$$

отсюда находим выражения для искомым δ и γ :

$$\begin{aligned} \delta &= \left\{ \sqrt{\frac{1}{N} \sum_{j=1}^N \left[\tau(X_j; \varepsilon, \lambda) - \frac{1}{N} \sum_{i=1}^N \tau(X_i; \varepsilon, \lambda) \right]^2} \right\}^{-1}, \\ \gamma &= \left\{ -\frac{1}{N} \sum_{j=1}^N \tau(X_j; \varepsilon, \lambda) \right\} \delta. \end{aligned} \quad (20)$$

Таким образом, доопределив параметры семейств кривых Джонсона на этапе обучения по выборочным данным, аппроксимации неизвестных условных функций плотности (17) используются в байесовых решающих правилах вида (1) [6].

В заключение отметим, что качество полученного решающего правила (1), определяемое эмпирическим риском (2), зависит не только от качества полученных аппроксимацией многомерных условных функций плотности (8), (17), но и от геометрии взаимного расположения точек обучающих последовательностей.

1. Репин В.Г., Тартаковский Г.П. Статистический синтез при априорной неопределенности и адаптации информационных систем. М.: Сов. радио, 1977. 432 с.
2. Пугачев В.С. Теория вероятностей и математическая статистика. М.: Наука, 1979. 496 с.
3. Ramberg J.S., Schmeiser V.W. // Communication of the Acm. 1972. V. 15. N 11. P. 987-990.
4. Кендалл М., Стьюарт А. Теория распределений. М.: Наука, 1966. 588 с.
5. Хан Г., Шапиро С. Статистические модели в инженерных задачах. М.: Мир, 1968. 395 с.
6. Серых А.П., Протасов К.Т., Боркун Ф.Я. Известия вузов. Нефть и газ. 1972. N 1. С. 3-9.

Институт оптики атмосферы
СО РАН, Томск

Поступила в редакцию
16 марта 1994 г.

К.Т. Protasov. Parametrization of the Probabilistic Distributions in Image Recognition Based on Normalizing Transformations.

In this paper we present a solution of a problem on reconstruction of multidimensional conditional functions of density of classes based on normalizing transformations, satisfying two conditions, within the framework of Bayes approach to construction of deciding rules for identification of images in the space of definitions of large dimensionality. The first condition is that the estimantes of the distributions of individual features of images should agree with the actual single-dimensional distributions quite reliably for practical purposes. The second condition is that the aproximating distribution should describe the statistical relations between the components of observation vectors.