

МЕТОД ФОРМИРОВАНИЯ ПРИЗНАКОВОГО ПРОСТРАНСТВА И КЛАССИФИКАЦИИ СОСТОЯНИЙ СЛОЖНЫХ ОБЪЕКТОВ

А.В. Харченко

(представил д.ф.-м.н., проф. А.И. Бых)

В статье предлагается метод распознавания состояний сложных объектов, основанный на использовании обучающей выборки, представленной в виде эмпирической таблицы. Предложен метод ранжирования элементов таблицы с целью выделения наиболее значимых признаков. Получены количественные оценки вероятности нахождения объекта в одном из состояний.

Основой классификации состояний сложных объектов являются измерения, позволяющие с различной степенью точности оценить множество параметров (признаков). В качестве признакового пространства выбирают систему потенциально полезных признаков V_i , где $i=1, 2, \dots, L$, по которым производится распознавание. Затем составляется таблица, каждая строка которой характеризует значения измеренных показателей $z_{(k,i)}$ для объектов O_k , где $k=1, 2, \dots, K$ (табл. 1). Данная таблица является неклассифицированной выборкой.

Таблица 1

Значения параметров, характеризующих состояния объектов

		Показатели								
		V_1	V_2	V_3	V_4	V_5	...	V_i	...	V_L
Объекты	O_1	$z_{(1,1)}$	$z_{(1,2)}$	$z_{(1,3)}$	$z_{(1,4)}$	$z_{(1,5)}$...	$z_{(1,i)}$...	$z_{(1,L)}$
	O_2	$z_{(2,1)}$	$z_{(2,2)}$	$z_{(2,3)}$	$z_{(2,4)}$	$z_{(2,5)}$...	$z_{(2,i)}$...	$z_{(2,L)}$

	O_k	$z_{(k,1)}$	$z_{(k,2)}$	$z_{(k,3)}$	$z_{(k,4)}$	$z_{(k,5)}$...	$z_{(k,i)}$...	$z_{(k,L)}$

	O_K	$z_{(K,1)}$	$z_{(K,2)}$	$z_{(K,3)}$	$z_{(K,4)}$	$z_{(K,5)}$...	$z_{(K,i)}$...	$z_{(K,L)}$

Используя метод экспертных оценок можно сгруппировать объекты по их состояниям, т.е. создать классифицированную выборку, представляющую собой таблицу, в которой каждому состоянию соответствует группа объектов (табл. 2). Выборочное среднее признаков, соответствующих i -му показателю, определяются по формуле

$$m_{\Omega_A}(V_i) = \frac{1}{K_A} \sum_{j=1}^{K_A} Z_{\Omega_A}(i, j), \quad (1)$$

где K_A – количество объектов в подмножестве, обладающих свойством Ω_A , а соответствующее нормированное к математическому ожиданию выборочное значение среднего квадрата отклонения (СКО) равно

$$\sigma_{\Omega_A}(V_i) = \frac{1}{m_{\Omega_A}(V_i)} \sqrt{\left[\frac{1}{K_A} \sum_{j=1}^{K_A} (z_{\Omega_A}(i,j) - m_{\Omega_A}(V_i))^2 \right]}. \quad (2)$$

Таблица 2
Значения показателей, характеризующих состояния объектов

		<i>Показатели</i>						
		V_1	V_2	...	V_i	...	V_L	
Состояния	Ω_A	O_{1A}	$z_{\Omega_A}(1, 1)$	$z_{\Omega_A}(1, 2)$...	$z_{\Omega_A}(1, i)$...	$z_{\Omega_A}(1, L)$
	
		O_{KA}	$z_{\Omega_A}(KA, 1)$	$z_{\Omega_A}(KA, 2)$...	$z_{\Omega_A}(KA, i)$...	$z_{\Omega_A}(KA, L)$
	Ω_B	O_{1B}	$z_{\Omega_B}(1, 1)$	$z_{\Omega_B}(1, 2)$...	$z_{\Omega_B}(1, i)$...	$z_{\Omega_B}(1, L)$
	
		O_{KB}	$z_{\Omega_B}(KA, 1)$	$z_{\Omega_B}(KA, 2)$...	$z_{\Omega_B}(KA, i)$...	$z_{\Omega_B}(KA, L)$
	Ω_C	O_{1C}	$z_{\Omega_C}(1, 1)$	$z_{\Omega_C}(1, 2)$...	$z_{\Omega_C}(1, i)$...	$z_{\Omega_C}(1, L)$
	
		O_{KC}	$z_{\Omega_C}(KA, 1)$	$z_{\Omega_C}(KA, 2)$...	$z_{\Omega_C}(KA, i)$...	$z_{\Omega_C}(KA, L)$

Используя формулы (1) и (2), а также данные табл. 2, можно охарактеризовать состояния объектов указанными выше статистическими параметрами, которые представлены в табл. 3.

Таблица 3
Значения выборочных средних и нормированных СКО для различных показателей и состояний

		Состояние							
		Ω_1		...	Ω_j		...	Ω_N	
Показатель	V_1	$m(1,1)$	$\sigma_H(1,1)$...	$m(1,j)$	$\sigma_H(1,j)$...	$m(1,N)$	$\sigma_H(1, N)$
	V_2	$m(2,1)$	$\sigma_H(2,1)$...	$m(2,j)$	$\sigma_H(2,j)$...	$m(2,N)$	$\sigma_H(2, N)$

	V_i	$m(i,1)$	$\sigma_H(i,1)$...	$m(i,j)$	$\sigma_H(i,j)$...	$m(i,N)$	$\sigma_H(i, N)$

	V_L	$m(L,1)$	$\sigma_H(L,1)$...	$m(L,j)$	$\sigma_H(L,j)$...	$m(L,N)$	$\sigma_H(L, N)$

Отличия средних значений показателей у объектов различных подмножеств являются признаками, позволяющими производить классифи-

кацию, а нормированные значения СКО характеризуют вариабельность соответствующего признака и относительную точность измерений. Учитывая, что в зависимости от конкретного состояния исследуемого объекта математические ожидания различных показателей могут отличаться от соответствующих математических ожиданий, полученных для неклассифицированной выборки, как в сторону увеличения, так и в сторону уменьшения, в качестве интегрального признака, характеризующего принадлежность объекта j -му состоянию, предлагается использовать функционал вида

$$Z_j = \begin{cases} \sum_{i=1}^{K_j} \frac{z_i - m_{ij}}{m_{ij}}, & \text{если } m_{ij} \geq m_{i,K}; \\ \sum_{i=1}^{K_j} \frac{z_i + m_{ij}}{m_{ij}}, & \text{если } m_{ij} < m_{i,K}, \end{cases} \quad (3)$$

где z_i – измеренное значение i -го показателя; m_{ij} – математическое ожидание i -го показателя группы объектов обладающих j -м признаком; K_j – количество объектов находящихся в классифицированной выборке, относящейся к j -му признаку; $m_{i,K}$ – математическое ожидание i -го показателя неклассифицированной выборки (табл. 2):

$$m_{i,K} = \frac{1}{K} \sum_{k=1}^K Z_{\Omega}(i,k), \quad (4)$$

где K – количество объектов находящихся в неклассифицированной выборке.

Для интегрального признака, представленного выражением (3), плотность распределения вероятности стремится к нормальной и имеет вид

$$w(Z_j) = \frac{1}{\sqrt{2\pi} \sum_{i=1}^L \left(\frac{\sigma_{ij}}{m_{i,K}} \right)^2} \exp \left(- \frac{\left(Z_j - \sum_{i=1}^L \frac{m_{ij} - m_{i,K}}{m_{i,K}} \right)^2}{2 \sum_{i=1}^L \left(\frac{\sigma_{ij}}{m_{i,K}} \right)^2} \right) \quad (5)$$

или

$$w_j(Z) = \frac{1}{\sqrt{2\pi} \Sigma_j} \exp \left(- \frac{(Z - M_j)^2}{2 \Sigma_j^2} \right), \quad (6)$$

где

$$M_j = \sum_{i=1}^L \left| \frac{m_{ij} - m_{i,K}}{m_{i,K}} \right| \quad (7)$$

– математическое ожидание значения интегрального признака соответствующего j -му состоянию объекта;

$$\Sigma_j^2 = \sum_{i=1}^L \left(\frac{\sigma_{ij}}{m_{i,K}} \right)^2 \quad (8)$$

– дисперсия значения интегрального признака соответствующего j -му состоянию объекта.

На рис. 1, в качестве примера, представлены результаты расчётов плотностей распределения интегрального признака соответствующего j -му состоянию объекта для различных форм тромбозмболии легочной артерии (ТЭЛА) у пациентов, выполненные по формуле (5) на основании данных, полученных в Институте общей и неотложной хирургии АМН Украины г. Харькова, на ультразвуковом аппарате Ultramark-9 и представленных в форме табл. 1. Вычисленные статистические характеристики интегральных признаков приведены в табл. 4.

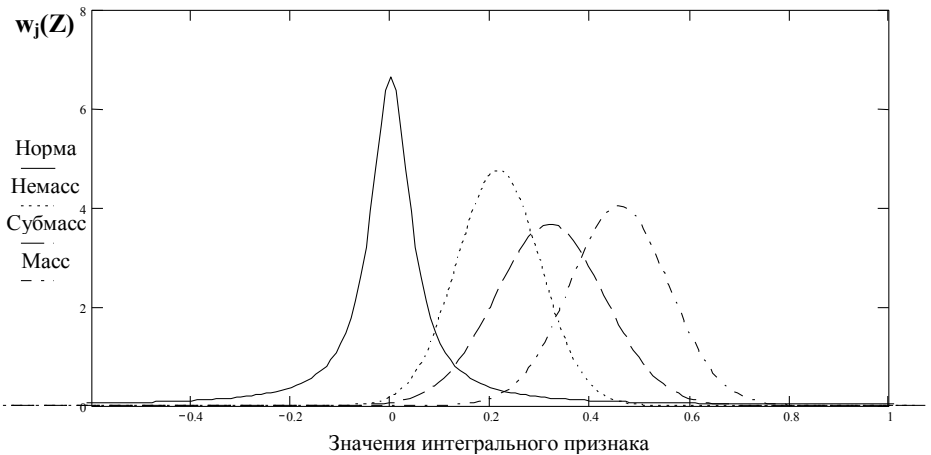


Рис. 1. Распределение плотности вероятности значений интегрального признака для различных форм ТЭЛА

Очевидно, что полученные распределения позволяют производить диагностику заболеваний по степени приближения значений интегрального признака исследуемого объекта к математическому ожиданию соответствующей плотности распределения (табл. 4).

Например, если измеренное значение интегрального признака у пациента равно $Z_{изм} = 0.07$, то с большей вероятностью можно предполо-

жить, что данный пациент не имеет ТЭЛА. Вероятность того, что у него ТЭЛА немассивной формы определяется выражением

$$P_{HM} = \frac{1}{\sqrt{2\pi} \Sigma_{HM}} \int_{-\infty}^{Z_{изм}} \exp\left(-\frac{(Z_j - M_{HM})^2}{2 \Sigma_{HM}^2}\right) dZ_j = P\left(\frac{Z_{изм} - M_{HM}}{\Sigma_{HM}}\right), \quad (9)$$

где $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$ - интеграл вероятности.

Подставляя в (9) $Z_{изм} = 0.07$ и взятые из табл. 4 значения $M_{HM} = 0.216$ и $\Sigma_{HM} = 0.084$, получим, что вероятность немассивной формы ТЭЛА у данного пациента равна $P_{HM} = F(-1.7381) = 1 - F(1.7381) = 0.0411$.

Заметим, что значимость различных показателей может существенно отличаться как между собой, так и между классами. В качестве универсального показателя значимости предлагается следующий функционал

$$R_i = \int_{-\infty}^{M_k} w_{ip}(Z) dz - \int_{-\infty}^{M_p} w_{ik}(Z) dz, \quad (10)$$

где $w_{ip}(Z)$ и $w_{ik}(Z)$ – плотности распределения значений i -х показателей для соответствующих p -му и k -му состояниям объекта; M_k и M_p – математические ожидания значений соответствующих показателей.

Таблица 4

Значения параметров распределения интегрального признака

Статистические характеристики интегрального признака	Состояния			
	Норма	Немассивная ТЭЛА (НМ)	Субмассивная ТЭЛА (СМ)	Массивная ТЭЛА (М)
МО	0,000	0,216	0,320	0,459
СКО	0,048	0,084	0,109	0,099

Наиболее значимый показатель имеет максимальное значение R_i на множестве показателей, определённых для различия p -го и k -го состояний. Предельно достижимым является $R_{i \max \max} = 1$.

Для примера, в соответствии с выражением (10) вычислены и представлены в табл. 5 значения параметров, характеризующие их значимость при диагностике ТЭЛА. Результаты ранжирования показателей по их значимости приведены в табл. 6.

Таблица 5

Значимость диагностических показателей
при дифференцировании различных форм ТЭЛА

Показатели	Нормы – Немасс	Немасс – Суб-масс	Субмасс – Масс
КДР ЛЖ	0,339	0,726	0,859
КДР ПЖ	0,991	0,246	0,181
$\frac{\text{КДР ПЖ}}{\text{КДР ЛЖ}}$	0,984	0,522	0,507
...
Е/А (ТТК)	0,097	0,774	0,563

Таблица 6

Ранжированные по значимости диагностические показатели
при дифференцировании различных форм ТЭЛА

Нормы – Немасс	Немасс – Субмасс	Субмасс - Масс
ПЖмл/ЛЖ мл	ср. давл	КДР ЛЖ
ПЖ м-л	Е/А (ТТК)	ПЖ м-л
КДР ПЖ	КДР ЛЖ	АР-АсТ
...
ФВ	АР-АсТ	ФВ

Очевидно, что на основании данных табл. 6 можно сделать выборку наиболее значимых показателей, которые в условиях ограничений на число измеряемых показателей дадут лучший результат при распознавании состояний человеческого организма.

Предложенный в статье метод распознавания, основанный на использовании введённого интегрального признака, включающего наиболее значимые показатели, позволяет автоматизировать принятие решения и оценить вероятность нахождения объекта в одном из допустимых состояний.

ЛИТЕРАТУРА

1. Нейлор К. Как построить экспертную систему. – М: Энергоатомиздат, 1991. – 286 с.

Поступила 30.07.2002

ХАРЧЕНКО Алексей Викторович, инженер отделения клинично-инструментальной и ультразвуковой диагностики Института общей и неотложной хирургии АМН Украины. В 2001 году окончил национальный аэрокосмический университет «ХАИ». Область научных интересов – вторичная обработка ультразвуковых изображений и программно-аппаратные методы выделения признаков сердечно-сосудистых заболеваний. E-mail: writeme@bigmir.net.