

УДК 681.518.25:537.533.9

ВИЗНАЧЕННЯ СИСТЕМИ НОРМОВАНИХ ДОПУСКІВ ПРИ РОЗПІЗНАВАННІ СПЕКТРОГРАМ

А.С.Краснопоясовський, доц.; С.О.Заговора, асп.; А.В.Черниш, асп.

Широке впровадження в електронній мікроскопії сучасних комп'ютерних технологій обробки спектрограм надає можливість шляхом застосування методів автоматичної класифікації зробити перехід від якісного аналізу спектрограм до кількісного. У межах методу функціонально-статистичних випробувань (МФСВ), який ґрунтується на оцінці інформаційної здатності системи розпізнавання, для формування ефективної навчальної вибірки (НВ) у процесі навчання здійснюється оптимізація в інформаційному сенсі контрольних допусків (КД) на ознаки розпізнавання (ОР) $\{d_{k,i}\}$, $i = \overline{1, N}$, де N - кількість ОР [1]. При цьому областю значень КД є відповідні поля нормованих допусків (НД) $\{d_{n,i}\}$, один із алгоритмів оптимізації яких розглянемо у даній роботі.

ПОСТАНОВКА ЗАДАЧІ

Один з можливих шляхів розв'язання задачі вибору оптимальної системи двобічних симетричних полів НД $\{a_{n,i}^*\}$ ґрунтується на гіпотезі: чим більша міжцентрова відстань для сусідніх (найближчих) класів розпізнавання, тим більша достовірність розпізнавання образу на екзамені, оскільки реалізації сусідніх класів будуть мати більше відмінних ОР. Тоді постановка задачі визначення НД у межах МФСВ полягає в наступному. Нехай $\{x_m\}$, $m = \overline{1, M}$, множина еталонних векторів (ЕВ) для відповідних класів розпізнавання $\{X_m^0\}$. Після розподілу множини $\{x_m\}$ на пари сусідніх векторів визначимо кодові відстані (КВ) між ними:

$$d_m = \sum_i^N (x_{m,i} \oplus x_{m+1,i}), \tag{1}$$

де $x_{m+1,i}$ - i -та координата сусіднього x_m ЕВ; \oplus - операція складання за модулем 2. Далі замість (1) будемо вживати скорочено: $d_m = d(x_m \oplus x_{m+1})$. Необхідно вибрати таку систему НД $\{\delta_{n,i}^*\}$, щоб середнє значення міжцентрової КВ для сусідніх класів розпізнавання

$$\bar{d}_m = \frac{1}{M} \sum_{m=1}^M d_m \tag{2}$$

було максимальним, тобто $\{d_{n,i}\} = \{\delta_{n,i}^*\}$, якщо $\bar{d}_m^* = \max_{\{m\}} \frac{1}{M} \sum_{m=1}^M d_m$.

Отже, критерій оптимізації (2) є медіаною міжцентрових КВ для сусідніх класів розпізнавання.

АЛГОРИТМ ОПТИМІЗАЦІЇ

Єдиним параметром оптимізації є величина поля допусків $|d_n|$, центром якого є дискретна підпірної реалізації спектра $\{e_{0,i}\}$. Алгоритм складається з проведення послідовних етапів, на кожному з яких формується при поточній системі допусків $\{d_{n,i}\}$ множина $\{x_m\}$, здійснюється її розподіл на пари сусідніх ЄВ за умови мінімальної КВ між ними, обчислюється за формулою (2) медіана міжцентрових КВ для сусідніх класів, і здійснюється наступний крок ітераційної процедури пошуку екстремуму \overline{d}_m^* . Вхідні дані: $\{e_{m,l,j} \mid m = \overline{1, M}, i = \overline{1, N}, j = \overline{1, n_{\min}}\}$ – масив дискретних спектрів, що аналізуються, де n_{\min} – мінімальний обсяг репрезентативної НВ, який обчислюється, наприклад, за алгоритмом [2]; d – змінна поля допусків; h – крок зміни поля допусків; S – зміна кроків ітерації. Вихідні дані: $\{\delta_{n,i}^*\}$, $\{x_{m,i}^*\}$ – система НД на ОР і масив ЄВ відповідно.

На підготовчому етапі обчислюється підпірна еталонна реалізація дискретних $\{e_{l,i}\}$ шляхом статистичного усереднення значень дискретних реалізацій спектра класу X_1^0 і задається вихідне значення d .

Для опису алгоритму достатньо розглянути схему алгоритму для одного S -го етапу, який складається з наступних кроків.

1 Формування масиву бінарних ЄВ $\{x_{m,i}\}$ за правилом

$$x_{m,i} = \begin{cases} 1, & \text{якщо } e_{0,i} - \delta \leq e_{m,i} \leq e_{0,i} + \delta, \\ 0, & \text{якщо інакше, або } e_{m,i} = 0. \end{cases} \quad (3)$$

2 Побудова матриці КВ (МКВ) $\|d_{k,l}\|$ за правилом

$$d_{k,l} = \begin{cases} \sum_{i=1}^N (x_{k,i} \oplus x_{l,i}), & \text{якщо } k \neq l, \\ N+1, & \text{якщо } k = l, \end{cases} \quad (4)$$

де k, l – змінні рядків і стовпців МКВ відповідно, $k, l = \overline{1, M}$.

3 Визначення мінімальних елементів рядків МКВ $\{\min d_{k,l}\}$.

4 Формування пар сусідніх ЄВ $R_k^{(2)} = \langle x_k, x_c \rangle$, де x_c сусідній x_k вектор, який має КВ $d_k(x_k \oplus x_c) = \min(d_{k,l})$.

5 Визначення міжцентрових КВ для сусідніх класів $d_k = d(x_k \oplus x_c)$.

6 Обчислення за формулою (2) медіани \overline{d}_m для множини $\{R_k^{(2)}\}$.

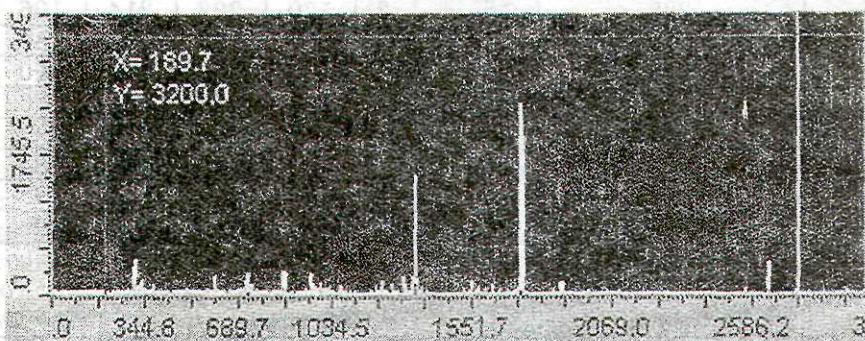
7 Пошук максимуму медіани \overline{d}_m^* шляхом порівняння поточного значення \overline{d}_m з попереднім \overline{d}_m' . Якщо $\overline{d}_m \geq \overline{d}_m'$, то $d+h$, і виконується крок 1, інакше $\overline{d}_m^* = \overline{d}_m'$ і $d^* = d-h$.

8 Формування вихідних даних: НД $|d_{n,i}^*| = (e_{0,i} + d^*) - (e_{0,i} - d^*) = 2d^*$, і множина ЄВ $\{x_{m,i}^*\}$, яка обчислюється за формулою (3) за умови $d=d^*$.

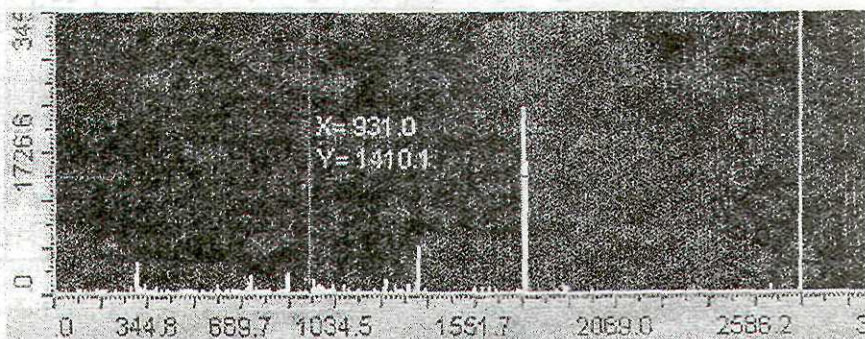
Оскільки ненульові дискрети спектрограм одного класу мають однакові детерміновано визначені енергетичні рівні, то формування повторних реалізацій спектрограм здійснюється тільки для цих рівнів.

ПРИКЛАД ОПТИМІЗАЦІЇ НОРМОВАНИХ ДОПУСКІВ

Реалізацію вищезазначеного алгоритму розглянемо на прикладі визначення оптимальних за критерієм (2) НД на значення дискрет спектрограм. На рис. 1 наведено дві реалізації спектрограми, які отримано на атомно-абсорбованому спектрофотометрі КАС-150 виробництва АТ СЕЛМІ (м. Суми, Україна) за двома зразками однакового хімічного складу. При цьому на осі абсцис відкладається довжина хвилі в нанометрах, а на осі ординат – значення струму на виході фотоелектронного підсилювача. Якщо поява дискрет в наведених реалізаціях відбувається при детермінованих значеннях довжини хвилі, то їх амплітуда має випадкове значення, що обумовлено перш за все розкидом геометричних параметрів зразків.



a)



б)

Рисунок 1

Таблиця 1 містить закруглені до цілих чисел значення 12 дискрет, які перевищують фоновий шум і отримані для семи спектрограм (перші дві з них наведено на рис.1а та рис.1б відповідно). В останньому стовпчику наведено дискрету усередненої реалізації e_0 . У таблиці 2 наведено бінарні ЕВ, які сформовано на базі таблиці 1 за правилом (3) при $d=40$. На базі таблиці 2 за правилом (4) сформовано МКВ, яку наведено в таблиці 3. За таблицею 3 визначимо пари сусідніх векторів: $R_1=\langle x_1, x_2 \rangle$, $R_2=\langle x_2, x_1 \rangle$, $R_3=\langle x_3, x_1 \rangle$, $R_4=\langle x_4, x_2 \rangle$, $R_5=\langle x_5, x_1 \rangle$, $R_6=\langle x_6, x_7 \rangle$, $R_7=\langle x_7, x_7 \rangle$. Відповідно КВ між сусідніми векторами

дорівнюють: $d(x_1 \oplus x_2)=2$, $d(x_2 \oplus x_1)=2$, $d(x_3 \oplus x_1)=3$, $d(x_4 \oplus x_2)=3$, $d(x_5 \oplus x_1)=2$, $d(x_6 \oplus x_7)=2$, $d(x_7 \oplus x_6)=2$. За формулою (2) медіана КВ дорівнює $d_m=2,29$. На рис.2 наведено залежності медіани $d_{cp} = \bar{d}_m$ від поля допусків δ .

Таблиця 1

	Реалізації спектрограми							e_0
	1	2	3	4	5	6	7	
1	466	408	440	382	418	434	450	444
2	174	116	152	156	142	192	144	154
3	288	313	290	326	294	230	284	277
4	291	288	220	310	294	348	200	294
5	288	168	182	170	222	214	196	206
6	174	170	205	158	202	194	182	184
7	1455	627	940	1120	1018	842	910	987
8	174	115	180	205	148	154	168	164
9	2328	2326	2410	2380	2520	2108	2562	2376
10	174	115	122	152	130	162	184	148
11	466	120	340	180	236	304	386	290
12	3520	3450	3210	3415	3505	3250	3282	3405

Таблиця 2

	Координати вектора											
	1	2	3	4	5	6	7	8	9	10	11	12
$X_m^{(1)}$	1	1	1	1	0	1	0	1	0	1	0	0
$X_m^{(2)}$	1	1	1	1	1	1	0	0	0	1	0	0
$X_m^{(3)}$	1	1	1	0	1	1	0	1	1	1	0	0
$X_m^{(4)}$	0	1	1	1	1	1	0	0	1	1	0	1
$x_m^{(5)}$	1	1	1	1	1	1	1	1	0	1	0	0
$x_m^{(6)}$	1	1	0	0	1	1	0	1	0	1	1	0
$x_m^{(7)}$	1	1	0	1	1	1	0	1	0	1	0	0

Таблиця 3

k	l						
	x_1	x_2	x_3	x_4	X_5	x_6	x_7
x_1	-	2	3	5	2	4	2
x_2	2	-	3	3	2	4	2
x_3	3	3	-	4	3	3	3
x_4	5	3	4	-	5	6	5
x_5	2	2	3	5	-	4	2
x_6	4	4	3	6	4	-	2
x_7	2	2	3	5	2	2	-

ПРИКІНЦЕВІ ПОЛОЖЕННЯ

Запропоновано визначати поле НД на значення дискрет спектрограм шляхом ітеративної процедури пошуку екстремуму критерію (2), який є медіаною міжцентрових КВ для пар сусідніх класів розпізнавання. Для задачі кількісної класифікації спектрограм поле НД доцільно визначати за реалізаціями базового класу X_1^0 , який характеризується усередненими показниками концентрації досліджуваного хімічного елемента. Аналіз рис.2 показує, що екстремальне значення поля НД $d^*=20$, яке забезпечує необхідні умови підвищення достовірності розпізнавання реалізацій класів, що утворюють алфавіт $\{X_m^0\}$, за рахунок максимізації міжцентрових відстаней між ЕВ сусідніх класів. При цьому робоча область значень критерію (2) визначається за умови формування ЕВ, що розрізняються. Так, при $d=5$ перша, друга і шоста реалізації



Рисунок 2

мають нульові ЕВ. З іншого боку, при збільшенні поля НД до $d=70$ з'являються однакові ЕВ для другої та четвертої реалізацій, що робить подальше збільшення поля НД недоцільним. При автоматичній класифікації спектрограм за МФСВ визначена за критерієм (2) система НД на ОР є вихідною і встановлює граничні значення відповідної системи КД, яка оптимізується в процесі навчання системи розпізнавання за інформаційним критерієм.

SUMMARY

The algorithm of definition of the system of normed bounds on discrete values of spectrogram obtained by spectrophotometry has been offered. The normed bounds are considered to be optimal in informational meaning. The median of code spacing on centers is used as a criterion of optimization for pairs of neighboring classes of recognition.

СПИСОК ЛІТЕРАТУРИ

1. Краснополюсовський А.С., Черниш А.В. Алгоритм навчання системи розпізнавання за методом функціонально-статистичних випробувань // Вісник СумДУ, 1998.- №1 (9).- С.89-94.
2. Марченко В.В., Краснополюсовський А.С. Формування репрезентативної навчальної виборки для систем контролю та діагностування // Обробка сигналів і зображень та розпізнавання образів: Праці Третьої всеукраїнської міжнародної конференції. - Київ, 1996.- С.107-109.

Надійшла до редколегії 25 березня 1999 р.