УДК 681.518:004.93.1

ВПЛИВ ПОТУЖНОСТІ СЛОВНИКА ОЗНАК РОЗПІЗНАВАННЯ НА СТРУКТУРУ БІНАРНОГО ПРОСТОРУ ОЗНАК

І. В. Шелехов, канд. техн. наук,
Сумський державний університет
kras@id.sumdu.edu.ua

Як критерій оптимізації процесу навчання системи підтримки прийняття рішень інформаційно-екстремальна інтелектуальна (ІЕІ) технологія застосовує статистичні інформаційні критерії функціональної ефективності, які є природною мірою різноманітності (або схожості) класів розпізнавання і одночасно функціоналом асимптотичних точнісних характеристик [1]. При цьому достовірність класифікатора залежить від геометричних параметрів контейнерів класів розпізнавання. В бінарному просторі наближенням гіперсферичного контейнера є гіперкуб. З метою узагальнення та зручності побудови такого контейнера припустимо існування псевдогіперсфери, яка описує гіперкуб, тобто містить усі його вершини. Це дозволяє в подальшому розглядати такі параметри оптимізації контейнера в радіальному базисі простору Хеммінга, як еталонний вектор, $x_m \in X_m^o$, вершина якого визначає геометричний центр контейнера $K_m^o \in X_m^o$, і радіус псевдосферичного контейнера, який визначається за формулою $d_m = \sum_{i=1}^{N}\left(x_{m,i} \oplus \lambda_i\right)$, де $x_{m,i}$ – i-та координата двійкового еталонного вектора x_m; λ_i – i-та координата вектора λ, вершини якого належать поверхні контейнера $K_m^o \in X_m^o$; $\oplus$ – операція додавання за модулем 2.

Розглянемо вплив на ефективність навчання потужності словника ознак. Нехай d_0^*, d_1^* – оптимальні радіуси контейнерів класів X_0^o та X_1^o відповідно, а $d_c = d\left(x_0^* \oplus x_1^*\right)$ – кодова відстань між їх центрами – еталонними векторами $x_0^* \in X_0^o$ і $x_1^* \in X_1^o$ відповідно. Враховуючи особливості бінарного простору Хеммінга, визначимо такі припущення:

1) потужність бінарного простору Хеммінга для словника ознак розпізнавання $\Sigma^{|N|}$ дорівнює 2^N;

2) кількість двійкових реалізацій на кодовій відстані d ($0 \le d \le N$) від двійкового вектора x дорівнює

$$B_d^N(x) = C_N^d = \frac{N!}{d!(N-d)!};$$

3) кількість бінарних реалізацій, що належить довільному контейнеру класу X_K^o з радіусом d_K ($0 \le d_K \le N$), дорівнює

$$B_{d,K}^N = \sum_{i=0}^{d_K} B_N^i = \sum_{i=0}^{d_K} \frac{N!}{i!(N-i)!}.$$

У випадку чіткого розбиття $\Re^{|M|}$ для M класів, тобто при $d_0^* + d_1^* < d_c$, кількість відповідних реалізацій, що належить контейнерам класів X_0^{O} і X_1^{O}, дорівнює

$$B_{\Re^{|m|}}^N = \sum_{k=0}^{m} B_{d,k}^N = \sum_{k=0}^{m}\sum_{i=0}^{d_k^*} \frac{N!}{i!(N-i)!},$$

а кількість реалізацій, що знаходяться поза межами цих контейнерів, дорівнює $B_{\overline{\Re}}^{N} = 2^{N} - B_{\Re}^{N}$.

Для загального випадку нечіткого розбиття $\tilde{\Re}^{|2|}$ для двох класів обчислення значення $B_{\tilde{\Re}^{|2|}}^{N}$ ускладнюється через наявність області простору $K_0^o \cap K_1^o$, де контейнери класів X_0^o і X_1° перетинаються. Для дослідження розподілу реалізацій образів визначимо такі припущення:

1) кількість бінарних реалізацій, що знаходяться на відстані d_0 від двійкового вектора x_0 та d_1 від двійкового вектора x_1, дорівнює 0, якщо $d_0 + d_1 < d(x_0 \oplus x_1)$;

2) кількість бінарних реалізацій, що знаходяться на відстані d_0 від вектора x_0 та d_1 від вектора x_1, дорівнює 0, якщо $d_1 = |d(x_0 \oplus x_1) - d_0| + 2p + 1$, де $p = 0,1,2,...$;

3) кількість двійкових реалізацій, що знаходяться на відстані d_0 від вектора x_0 та відстані d_1 від вектора x_1, якщо $d_1 = d(x_0 \oplus x_1) - d_0$ та $d_0 \le d(x_0 \oplus x_1)$, дорівнює

$$n = \frac{d(x_0 \oplus x_1)!}{d_0!(d(x_0 \oplus x_1) - d_0)!};$$

4) кількість двійкових реалізацій, що знаходяться на кодовій відстані d_0 від двійкового вектора x_0 та d_1 від двійкового вектора x_1, дорівнює

$$n = \frac{d(x_0 \oplus x_1)!}{(d_0 - p)!(d(x_0 \oplus x_1) - d_0 + p)!} \cdot \frac{(N - d(x_0 \oplus x_1))!}{(N - d(x_0 \oplus x_1) - p)!\,p!},$$

де $p = \dfrac{d_1 + d_0 - d(x_0 \oplus x_1)}{2}$.

На рис. 1 показано структуру десятивимірного бінарного простру, в якому відтворено контейнери двох класів X_0^o та X_1^o, що перетинаються. При цьому діаметр заповнених кіл відповідає кількості двійкових реалізацій, що характеризуються відповідними відстанями від центрів зображених контейнерів.

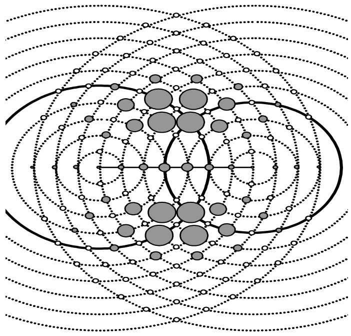

Рис. 1. Структура простору Хеммінга при побудові контейнерів за принципом “найближчого сусіда”

Таким чином, деталізований аналіз структури розбиття класів у просторі Хеммінга свідчить про імплікативність і симетричність розподілу векторів-реалізацій класів у відповідних контейнерах, побудованих в радіальному базисі простору ознак.

1. Довбиш А.С. Основи проектування інтелектуальних систем: навчальний посібник. – Суми: Видавництво СумДУ, 2009.– 171 с.