

ІДЕНТИФІКАЦІЯ ГОЛОСУ

О.А. Якушев, аспірант

Сумський державний університет, м. Суми

Рассматривается синтез системы идентификации голоса в рамках информационно-экстремальной интеллектуальной технологии, которая основана на максимизации информационной возможности системы путем введения в процесс её обучения дополнительных информационных ограничений.

Розглядається синтез системи ідентифікації голосу у рамках інформаційно-екстремальної інтелектуальної технології, що ґрунтується на максимізації інформаційної спроможності системи шляхом введення в процес її навчання додаткових інформаційних обмежень.

ВСТУП

Розроблення високодостовірних і оперативних алгоритмів ідентифікації голосу залишається актуальним, наприклад, у зв'язку з підвищенням вимог до інформаційної захищеності систем керування з обмеженим доступом, складних робототехнічних систем і комплексів та інше. Існує два основні підходи до вирішення проблеми розв'язання розпізнавання голосу:

- структурний, що ґрунтується на синтезі образів [1-2];
- підхід, орієнтований на використання навчальної вибірки [2-3].

Основним недоліком структурного підходу є необхідність на етапі екзамену синтезу образу, що розпізнається, з метою його порівняння з еталонним образом, сформованим на етапі навчання системи. Це створює суттєві ускладнення для підвищення оперативності ідентифікації та розпізнавання голосу.

У рамках другого підходу обов'язковим є наявність багатовимірної навчальної матриці, результатом якого є побудова вирішальних правил, що застосовується в режимі екзамену для безпосереднього розпізнавання голосу. Перевагою такого підходу є можливість прийняття високоточних рішень або результатів розпізнавання. При цьому центральною задачею є побудова на етапі навчання деяким способом оптимального тут і далі в інформаційному розумінні розбиття простору ознак на класи.

Існуючі методи розпізнавання (LIRA, проект розроблювачів "СТЭЛ - Комп'ютерні Системи" та інші [2-3]) мають в основному модельний характер, оскільки ігнорують апріорне нечітке розбиття класів розпізнавання до задач практики. Є методи, що створено або розроблено у рамках інформаційно-екстремальної інтелектуальної технології (ІЕІТ) [4], що ґрунтується на максимізації інформаційної спроможності системи розпізнавання, що навчається, шляхом введення додаткових обмежень в процесі оптимізації просторово-часових параметрів функціонування.

ПОСТАНОВКА ЗАВДАННЯ

Метою створення системи ідентифікації голосу користувача є розроблення інформаційного та програмного забезпечення системи розпізнавання, що навчається, у рамках ІЕІТ [4-9].

Розглянемо таку постановку задачі ідентифікації голосу у рамках ІЕІТ.

Нехай дано алфавіт класів розпізнавання $\{X_m^o \mid m = \overline{1, M}\}$, де M — кількість класів, серед яких клас X_1^o характеризує акустичні властивості голосу користувача, та навчальна матриця

$\| y_{m,i}^{(j)} \mid i = 1, N, j = 1, n \|$, де N – кількість ознак розпізнавання; n – кількість реалізацій образу. Відомий вектор параметрів функціонування системи розпізнавання, який у загальному випадку має таку структуру:

$$g = \langle g_1, \dots, g_{\xi_1}, \dots, g_{\Xi_1}, f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2} \rangle, \quad \Xi_1 + \Xi_2 = \Xi,$$

де $\langle g_1, \dots, g_{\xi_1}, \dots, g_{\Xi_1} \rangle$ – генотипні параметри функціонування нечіткого регулятора, які впливають на параметри розподілу реалізацій образу;

$\langle f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2} \rangle$ – фенотипні параметри функціонування нечіткого регулятора, які прямо впливають на геометрію контейнера класу розпізнавання.

При цьому відомі обмеження на відповідні параметри функціонування:

$$R_{\xi_1}(g_1, \dots, g_{\xi_1}, \dots, g_{\Xi_1}) \leq 0; \quad R_{\xi_2}(f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2}) \leq 0.$$

Треба на етапі навчання у рамках ІЕІТ побудувати в радіальному базисі бінарного простору ознак розпізнавання оптимальне в інформаційному розумінні розбиття класів розпізнавання за таких умов:

- 1) $(\forall X_m^o \in \tilde{\mathfrak{R}}^{|M|}) \quad [X_m^o \neq \emptyset]$;
- 2) $(\exists X_k^o \in \tilde{\mathfrak{R}}^{|M|}) \quad (\exists X_l^o \in \tilde{\mathfrak{R}}^{|M|}) \quad [X_k^o \neq X_l^o \rightarrow X_k^o \cap X_l^o \neq \emptyset]$;
- 3) $(\forall X_k^o \in \tilde{\mathfrak{R}}^{|M|}) \quad (\forall X_l^o \in \tilde{\mathfrak{R}}^{|M|}) \quad [X_k^o \neq X_l^o \rightarrow \text{Ker} X_k^o \cap \text{Ker} X_l^o = \emptyset]$;
- 4) $(\forall X_k^o \in \tilde{\mathfrak{R}}^{|M|}) \quad (\forall X_l^o \in \tilde{\mathfrak{R}}^{|M|})$
 $[X_k^o \neq X_l^o \rightarrow (d_k^* < d(x_k \oplus x_l)) \& (d_l^* < d(x_k \oplus x_l))];$
- 5) $\bigcup_{X_m^o \in \tilde{\mathfrak{R}}} X_m^o \subseteq \Omega_B$; $k \neq l$; $k, l, m = \overline{1, 2}$,

де d_k^* , d_l^* – оптимальні радіуси гіперсферичних контейнерів класів розпізнавання $K_k^o \in X_k^o$ і $K_l^o \in X_l^o$ відповідно.

При цьому інформаційний критерій функціональної ефективності (КФЕ) набуває максимуму в робочій області визначення його функції:

$$E^* = \max_G E_m,$$

де E_m – критерій функціональної ефективності навчання системи розпізнавати реалізації класу X_m^o ; G – робоча область КФЕ.

На етапі екзамену треба за побудованим на етапі навчання системи розпізнавання вирішальним правилом визначити належність реалізації

образу, що розпізнається, до відповідного класу розпізнавання із заданого алфавіту $\{X_m^0\}$.

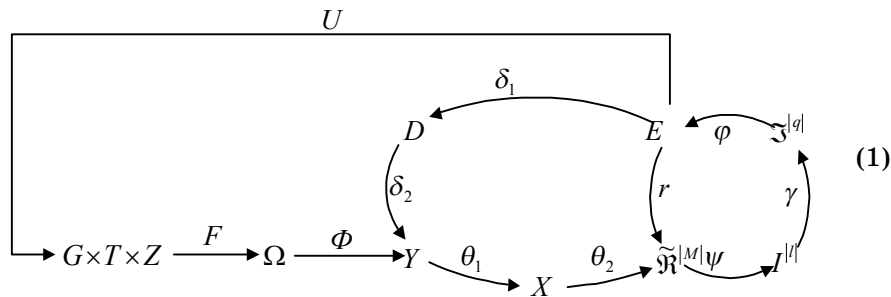
МАТЕМАТИЧНА МОДЕЛЬ

Розглянемо категорійні моделі навчання і екзамену в рамках ІЕІТ за апіорно класифікованими навчальними матрицями. Така модель повинна обов'язково передбачати вхідний математичний опис системи, який подамо у вигляді структури

$$\Delta = \langle G, T, \Omega, Z, Y, F, \Phi, \theta \rangle,$$

де G – простір вхідних сигналів (факторів); T – множина моментів часу зняття інформації; Ω – простір ознак розпізнавання; Z – простір можливих функціональних станів системи; Y – терм-множина сигналів після первинної обробки інформації (вбіркова множина); $F: G \times T \times \Omega \rightarrow Z$ – оператор формування простору Z ; $\Phi: G \times T \times \Omega \times Z \rightarrow Y$ – оператор формування вбіркової множини Y .

Введемо оператор θ побудови у загальному випадку нечіткого розбиття $\tilde{\mathfrak{R}}^{|M|}$ простору ознак на M класів розпізнавання. Нехай перевірка основної статистичної гіпотези про належність реалізацій навчальної матриці класу X_m^0 здійснюється оператором класифікації $\Psi: \tilde{\mathfrak{R}}^{|M|} \rightarrow I^{|l|}$, де $I^{|l|}$ – множина розв'язків (гіпотез), а l – кількість статистичних гіпотез. Таким чином, з урахуванням вищевведених операторів і множин математичну модель подамо у вигляді діаграми відображень множин, які застосовуються в процесі навчання:



У діаграмі (1) $\mathfrak{S}^{|q|}$ – множина точнісних характеристик процесу прийняття рішень, де $q = l^2$; E – множина значень КФЕ; D – система контрольних допусків на ознаки розпізнавання. Оператор U – регламентує процес навчання. У діаграмі правий контур операторів реалізує базовий алгоритм навчання за ІЕІТ, який обчислює КФЕ й оптимізує геометричні параметри контейнера відповідного класу, а контур $\theta \rightarrow \Psi \rightarrow \gamma \rightarrow \phi \rightarrow \delta_1 \rightarrow \delta_2 \rightarrow \theta$ оптимізує контрольні допуски на ознаки розпізнавання.

АЛГОРИТМ НАВЧАННЯ

Оптимізацію СКД будемо здійснювати за паралельним алгоритмом, при якому контрольні допуски оптимізуються для всіх ознак одночасно [4]. За цим алгоритмом оптимізуються параметри контейнерів класів розпізнавання за умови визначення оптимальних контрольних допусків для базового (еталонного) класу X_1^0 . При цьому центрування контейнерів здійснюється в процесі навчання шляхом знаходження еталонних

векторів-реалізацій образу, вершини яких визначають геометричні центри класів розпізнавання.

Як критерії оптимізації параметрів функціонування системи розпізнавання у рамках ІЕІТ можуть застосовуватися статистичні інформаційні критерії, які є функціоналами від точнісних характеристик класифікатора. Наприклад, модифікована міра Кульбака для двохальтернативної системи оцінок процесу навчання розпізнаванню реалізацій класу X_m^o за ІЕІТ при рівноймовірних апіорних гіпотезах має вигляд [4]:

$$\begin{aligned} J_m^{(k)} &= 0,5 \log_2 \left(\frac{D_1^{(k)} + D_2^{(k)}}{\alpha^{(k)} + \beta^{(k)}} \right) \left[(D_1^{(k)} + D_2^{(k)}) - (\alpha^{(k)} + \beta^{(k)}) \right] = \\ &= \log_2 \left(\frac{2 - (\alpha^{(k)} + \beta^{(k)})}{\alpha^{(k)} + \beta^{(k)}} \right) \left[1 - (\alpha^{(k)} + \beta^{(k)}) \right], \end{aligned} \quad (2)$$

де $D_1^{(k)}$, $D_2^{(k)}$, $\alpha^{(k)}$, $\beta^{(k)}$ – точнісні характеристики процесу навчання: перша та друга достовірності, помилки першого та другого роду відповідно, отримані на k -му кроці навчання.

Вхідною інформацією для навчання системи розпізнавання є навчальна матриця $\| y_{m,i}^{(j)} \mid m = \overline{1, M}; i = \overline{1, N}, j = \overline{1, n^*} \|$, система полів нормованих допусків $\{\delta_{H,i}\}$, яка задає область значень контрольних допусків $\{\delta_{K,i}\}$.

Нехай алфавіт характеризується двома класами:

- 1) клас X_1^o , який характеризує голос «хазяїна»;
- 2) клас X_2^o , який характеризує чужий голос.

Нехай кожний клас буде складатися з 40 реалізацій, кожна з яких визначає сказану фразу. Кожна реалізація складається з 100 ознак. Обробка з початкового акустичного сигналу відбувалася в такий спосіб:

- 1) нормалізація звукового сигналу [10], зведення усіх сигналів до однієї гучності;
- 2) безпосереднє перетворення, квантування в децибелах за амплітудою на декілька рівнів і обчислення кількості протилежних екстремумів на кожному кроці квантування;
- 3) нормування отриманих значень, тобто кожне значення, ділиться на загальну кількість максимумів та мінімумів;
- 4) помноження нормованих значень на деяке фіксоване значення.

РЕЗУЛЬТАТИ ФІЗИЧНОГО МОДЕЛЮВАННЯ

Результати оптимізації СКД у процесі навчання системи ідентифікації голосу користувача ілюструє рис. 1.

На рис. 1 і далі темною ділянкою графіка позначено робочу (допустиму) область визначення функції (2), в якій $D_{1,1}^{(k)}$ і $D_{1,2}^{(k)}$ більше 0,5. Аналіз рис. 1 показує, що оптимальне значення параметра δ належить інтервалу [18;21]. Визначення конкретного оптимального значення здійснюється за мінімальним значенням критерію нечіткої компактності векторів-реалізацій класу X_1^o на зазначеному вище інтервалі:

$$l_{1,d} = \frac{d_1^*}{d(x_1 \oplus x_2)} \rightarrow \min_{G_d},$$

де d_1^* – оптимальний радіус контейнера класу X_1^o ; $d(x_1 \oplus x_2)$ – кодова міжцентрова відстань класів X_1^o і X_2^o ; G_d – допустима область значень радіусів контейнера класу X_1^o : $d \in [0; d(x_1 \oplus x_2) - 1]$.

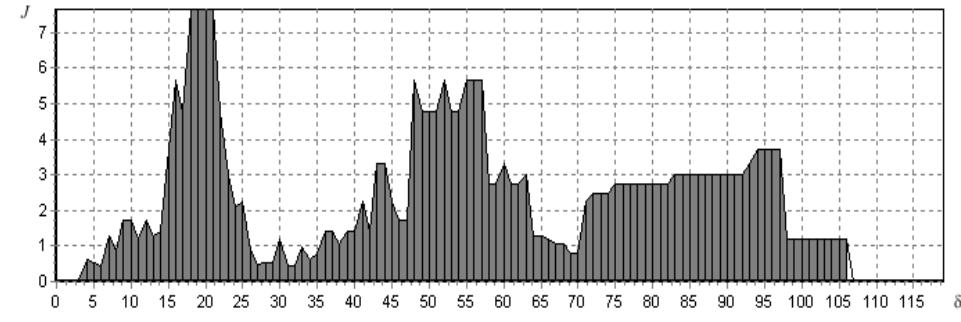


Рисунок 1 – Графік залежності інформаційного критерію від параметра поля допусків

Більш детальну інформацію про оптимізацію контрольних допусків наведено у табл. 1, де темною ділянкою позначено робочу (допустиму) область визначення функції критерію (2).

Таблиця 1 – Оптимізація контрольних допусків за паралельним алгоритмом

δ	J	D_1	β	d_m	d_c	l_δ
0	0	1	1	1	0	-
1	0,00178	0,575	0,6	14	1	14
2	0,04486	0,6	0,475	19	9	2,111111
3	0,02866	0,55	0,45	18	11	1,636364
4	0,62211	0,95	0,5	22	17	1,294118
5	0,55018	0,875	0,45	20	16	1,25
6	0,42189	0,775	0,4	18	14	1,285714
...
15	3,70619	1	0,1	11	13	0,846154
16	5,67999	0,975	0	8	11	0,727273
17	4,77825	0,95	0	8	12	0,666667
18	7,65105	1	0	9	12	0,75
19	7,65105	1	0	8	11	0,727273
20	7,65105	1	0	7	10	0,7
21	7,65105	1	0	7	10	0,7
22	4,77825	0,95	0	6	9	0,666667
23	3,00828	1	0,15	6	8	0,75
24	2,06924	1	0,25	6	7	0,857143
25	2,26702	1	0,225	6	6	1
26	0,87264	1	0,475	6	6	1

Продовження табл. 1

27	0,48352	0,875	0,475	5	6	0,833333
28	0,55018	0,875	0,45	5	6	0,833333
29	0,55018	0,5	0,075	3	5	0,6
30	1,18401	0,75	0,15	4	5	0,8
...
180	0	1	1	1	0	-

ВИСНОВКИ

У процесі навчання у рамках ІЕІТ побудовано оптимальний в інформаційному сенсі гіперсферичний класифікатор, який дозволяє ідентифікувати голос користувача з близькою до одиниці достовірністю.

У перспективі на базі розробленого алгоритму ідентифікації голосу планується створення алгоритму розпізнавання команд за умови розширення словника ознак та додаткової оптимізації інших параметрів функціонування системи розпізнавання

SUMMARY

VOICE IDENTIFICATION

O.A. Yakushev

Sumy State University

The synthesis of the voice authentication system is examined within the bounds of information-extreme intelligence technology which is based on informative possibility maximization of the system by introduction in the process of its learning additional informative limitations.

СПИСОК ЛІТЕРАТУРИ

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – Киев: Наукова думка, 1987. -262 с.
2. Информационный портал речевых технологий <http://art.bdk.com.ru/govor/>
3. Касаткина Л.М., Лукович В.В., Пилипенко В.В. Распознавание личности по голосу с помощью классификатора LIRA// Управляющие системы и машины.
4. Краснопопосовський А. С. Інформаційний синтез інтелектуальних систем керування: Підхід, що ґрунтується на методі функціонально-статистичних випробувань.–Суми: Видавництво СумДУ, 2004. – 261 с.
5. Васильев В. И. Распознающие системы: Справочник. - 2-е изд., перераб. и доп. – Киев: Наукова думка, 1983. – 422 с.
6. Васильев В.И., Шевченко А.И. Формирование и опознавание образов. – Донецк: ДонГИИИ, 2000. – 360 с.
7. Ивахненко А. Г. О принципах построения обучающихся систем управления сложными процессами. – М.: Наука, 1970. – 252 с.
8. Цыпкин Я.З. Основы теории обучающихся систем. – М.: Наука, 1970. – 251 с.
9. Аркадьев А.И., Браверман Э.М. Обучение машины классификации. – М.: Наука. Главная редакция физико-математической литературы, 1971. – 192 с.
10. Sadaoki Furui, M. Mohan Sondhi Advances in speech signal processing. – New York, 1992.

Надійшла до редакції 11 лютого 2009 р.