

ОПТИМІЗАЦІЯ КРОКУ КВАНТУВАННЯ АКУСТИЧНИХ СИГНАЛІВ ПРИ РОЗПІЗНАВАННІ ГОЛОСОВИХ КОМАНД

А.С. Довбиш, д-р техн. наук, професор;

О.А. Якушев, аспірант

Сумський державний університет, м. Суми

Розглядається процес оптимізації кроку квантування в часі акустичних сигналів при розпізнаванні голосових команд у рамках інформаційно-екстремальної інтелектуальної технології, що ґрунтується на максимізації інформаційної спроможності системи шляхом введення в процесі її навчання додаткових інформаційних обмежень.

Ключові слова: *ІЕІТ, крок квантування, класи розпізнавання, критерій функціональної ефективності Кульбака, звуковий сигнал, робоча область, алгоритм навчання.*

Рассматривается процесс оптимизации шага квантования во времени акустических сигналов при распознавании голосовых команд в рамках информационно-экстремальной интеллектуальной технологии, которая основана на максимизации информационной возможности системы, вводя в процесс её обучения дополнительных информационных ограничений.

Ключевые слова: *ИЭИТ, шаг квантования, классы распознавания, критерий функциональной эффективности Кульбака, звуковой сигнал, рабочая область, алгоритм обучения.*

ВСТУП

Функціональна ефективність системи розпізнавання мови визначається багатьма параметрами, серед яких важливу роль відіграють точність розпізнавання, стійкість системи до шумів, ступінь залежності від диктора та акустичної апаратури та інше. Тому проблема розпізнавання голосу все ще залишається невирішеною через її складність. Для розпізнавання голосових команд на практиці застосовуються методи, побудовані на основі марковських процесів [1,2] і штучних нейромереж [3,4], але вони не забезпечують достатньо високої достовірності розпізнавання через відсутність оптимізації параметрів функціонування системи розпізнавання за загальним інформаційним критерієм. Оскільки значний вплив на функціонування таких систем здійснюють параметри початкового оброблення сигналу, то їх оптимізація дозволяє підвищити точність та пристосованість систем розпізнавання акустичних сигналів. Одним із таких важливих параметрів є крок квантування в часі (дискретизації) акустичного сигналу. У статті розглядається алгоритм оптимізації кроку квантування в часі акустичного сигналу, розроблений за інформаційно-екстремальною інтелектуальною технологією (ІЕІ-технологія), що ґрунтується на максимізації інформаційної спроможності системи шляхом введення в процесі навчання додаткових інформаційних обмежень [5].

ПОСТАНОВКА ЗАДАЧІ

Один із шляхів підвищення достовірності автоматичної класифікації полягає в адаптації математичного опису на вході системи розпізнавання, що навчається, з метою досягнення на екзаміні достовірності розпізнавання, близької до асимптотичної, що визначається функціональною ефективністю навчання системи. Розглянемо постановку задачі стосовно гіперсферичного класифікатора, який будується в процесі навчання системи шляхом оптимізації контейнерів класів розпізнавання, що відновлюються в радіальному базисі простору ознак.

Нехай дано алфавіт класів розпізнавання $\{X_m^0 \mid m = \overline{1, M}\}$ і вектор параметрів функціонування системи розпізнавання $g_m = \{d_m, x_m, \delta, \tau\}$, де d_m – радіус контейнера $K_m^0 \in X_m^0$; x_m – двійковий еталонний вектор-реалізація класу X_m^0 , $x_m \in \Omega_B$; δ – параметр контрольного поля допусків $\delta \leq \delta_H / 2$, де δ_H – нормоване (експлуатаційне) поле допусків; τ – крок квантування вхідної реалізації.

Потрібно визначити оптимальні значення параметра τ^* – кроку квантування в часі реалізацій акустичного образу:

$$\tau^* = \arg \max_{g \in G} E_m,$$

де E_m – критерій функціональної ефективності навчання системи розпізнавати реалізації класу X_m^0 ; G – область допустимих значень параметрів функціонування.

МАТЕМАТИЧНА МОДЕЛЬ

Розглянемо категорійну модель навчання системи розпізнавання в рамках ІЕІ-технології за апріорно класифікованими навчальними матрицями [5]. Така модель повинна включати вхідний математичний опис системи, який подамо у вигляді структури

$$\Delta_B = \langle G, T, \Omega, Z, Y; F, \Phi \rangle,$$

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

Оскільки процес навчання системи полягає в побудові оптимального розбиття простору ознак Ω на класи, то введемо оператор θ побудови у загальному випадку нечіткого розбиття $\tilde{\mathfrak{H}}^{|M|}$ на M класів розпізнавання. Нехай перевірка основної статистичної гіпотези про належність реалізацій навчальної матриці класу X_m^0 здійснюється оператором класифікації $\Psi: \tilde{\mathfrak{H}}^{|M|} \rightarrow I^{|l|}$, де $I^{|l|}$ – множина рішень (гіпотез), а l – її потужність. На рис.1 наведено діаграму відображення множин, задіяних у процесі навчання системи розпізнавання з оптимізацією кроку дискретизації акустичного сигналу

На рис.1 контур оптимізації кроку квантування в часі замикається оператором $\tau_1: E \mapsto \check{T}$, де \check{T} – множина значень кроку квантування, і оператором $\tau_2: \check{T} \mapsto T$, який змінює інтервали часу зняття інформації.

Оптимізація кроку квантування в часі реалізацій образу в рамках інформаційно-екстремального алгоритму навчання згідно з діаграмою (рис. 1) здійснюється за таким ітераційним алгоритмом:

$$\tau^* = \arg \max_{G_\tau} \{ \max_{G_\delta} \{ \max_{G_d} E_1 \} \}, \quad (1)$$

де G_τ – допустима область значень параметра τ ; G_δ – допустима область значень параметра δ ; G_d – допустима область значень параметра d_1 .

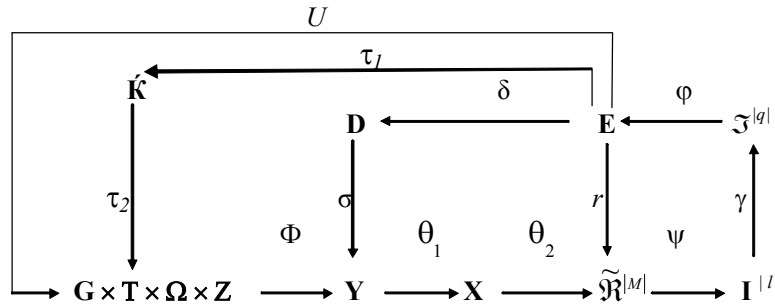


Рисунок 1 - Модель оптимізації кроку дискретизації акустичного сигналу

АЛГОРИТМ НАВЧАННЯ

Як критерії оптимізації параметрів функціонування системи розпізнавання у рамках ІЕІ-технології можуть застосовуватися статистичні інформаційні критерії, які є функціоналами від точнісних характеристик класифікатора. Наприклад, інформаційна міра Кульбака для двохальтернативної системи оцінок процесу навчання розпізнавання реалізацій класу X_m^o при рівноймовірних гіпотезах $p(\gamma_1) = p(\gamma_2) = 0,5$ має вигляд [5]

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

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.

Оптимізація кроку дискретизації здійснюється за таким алгоритмом:

1. Формування посуми кроків квантування $\tau := \tau + \Delta\tau$, де $\tau(0) := \tau_0$, який попередньо набуває деякого початкового значення τ_0 ;
2. Формування для базового класу X_1^0 навчальної матриці $\|y_{1,i}^{(j)}\|$ для поточного кроку дискретизації τ , де $j = \overline{1, n}$ – змінна числа реалізацій образу, $i = \overline{1, N}$ – змінна числа ознак розпізнавання.
3. Формування еталонної реалізацій акустичного сигналу $y_1 \in X_1^0$ шляхом статистичного усереднення стовпців матриці $\|y_{1,i}^{(j)}\|$.
4. Формування бінарної навчальної матриці $\|x_{1,i}^{(j)}\|$ для класу X_1^0 за правилом

$$x_{1,i}^{(j)} = \begin{cases} 1, & \text{if } y_{1,i}^{(j)} \in \delta_{k,i}; \\ 0, & \text{if } \text{else} \end{cases}$$

де $\delta_{k,i}$ – поле контрольних допусків на i -ту ознаку розпізнавання.

5. Формування для класу X_2^0 навчальної матриці $\|y_{2,i}^{(j)}\|$.

6. Формування бінарної навчальної матриці $\|x_{2,i}^{(j)}\|$ для класу X_2^0 аналогічно кроку 3 за умови використання системи контрольних допусків базового класу X_1^0 .

7. Формування лічильника кроків навчання (збільшення радіусів контейнера): $d:=d+1$ ($d(0) = 0$).

8. Для базового класу X_1^0 з метою побудови оптимального в інформаційному розумінні гіперсферичного контейнера реалізується ітераційний алгоритм навчання системи розпізнавання з паралельною оптимізацією контрольних допусків на ознаки розпізнавання [5].

9. На кожному кроці навчання обчислюється значення критерію за формулою (2).

10. Здійснюється пошук глобального значення критерію (2) E_1^* в робочій (допустимій) області визначення його функції. Порівняння: якщо $E_1^*(\tau) \geq E_1^*(\tau + \Delta\tau)$, то виконується крок 1, інакше – крок 12.

11. Порівняння: якщо $\tau \leq \tau_{\max}$, де τ_{\max} – максимальний крок квантування акустичного сигналу, то виконується крок 1, інакше – крок 13.

12. Визначення оптимального значення кроку квантування τ^* в часі реалізацій образів i .

13. Зупин.

Таким чином, оптимізація кроку квантування в часі акустичних сигналів у рамках розглянутого алгоритму полягає в ітераційному наближенні в процесі навчання значення глобального максимуму інформаційного критерію (5), обчисленого в робочій області визначення його функції, до граничного, яке характеризується відсутністю помилок першого та другого роду.

1.1 ПРИКЛАД РЕАЛІЗАЦІЇ АЛГОРИТМУ

Оптимізація кроку квантування в часі акустичних сигналів здійснювалася для двох класів, які характеризували дві команди: «вперед» та «назад». На рис.2 наведено графік залежності КФЕ (2) від кроку квантування τ . При цьому одна градація кроку квантування дорівнює $\frac{1}{160}$ с.

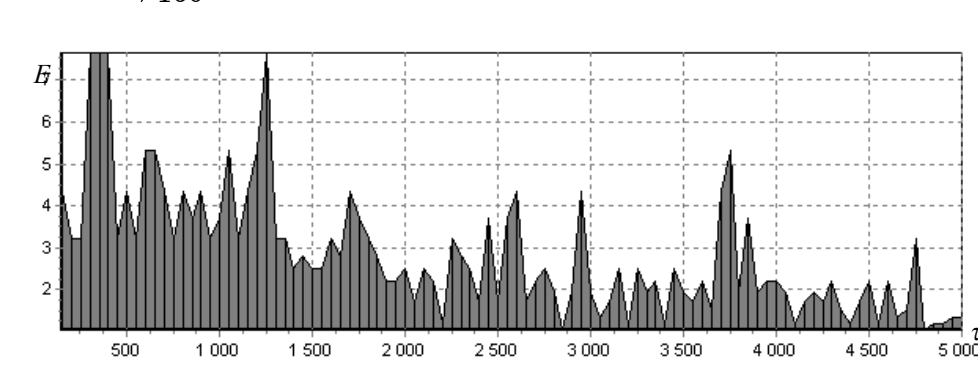


Рисунок 2 – Графік залежності КФЕ Кульбака від кроку квантування

Темною ділянкою на рис. 2 позначено робочу (допустиму) область визначення функції критерію (2). За оптимальне слід прийняти екстремальне значення кроку квантування, при якому інформаційний

критерій оптимізації (2) набуває максимального граничного значення. Аналіз рис. 2 показує, що оптимальний крок квантування належить множині значень $\tau^* = [300; 400] \cup \{500\}$. Визначення конкретного оптимального значення здійснюється за мінімальним значенням критерію нечіткої компактності векторів-реалізацій класу X_1^o на вказаному 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 : $d \in [0; d(x_1 \oplus x_2) - 1]$.

Більш детальна інформація наведена в табл. 1. У табл. 1 прийнято такі позначення: E – КФЕ (2), D_1 та D_2 – відповідно перша та друга достовірності, d_1 – радіус контейнера класу X_1^o , d_c – кодова відстань між центрами контейнерів класів X_1^o і X_2^o .

Таблиця 1 – Оптимізація радіуса контейнера для першого класу

| τ | J | D_1 | D_2 | d_1 | d_c | l_d |
|------------|-----------------|----------|----------|-----------|-----------|-----------------|
| 150 | 4,352871 | 0,966667 | 0,966667 | 7 | 10 | 0,7 |
| 200 | 3,215963 | 0,9 | 0,966667 | 5 | 7 | 0,714286 |
| 250 | 3,215963 | 0,933333 | 0,933333 | 11 | 14 | 0,785714 |
| 300 | 7,651052 | 1 | 1 | 14 | 17 | 0,823529 |
| 350 | 7,651052 | 1 | 1 | 10 | 13 | 0,769231 |
| 400 | 7,651052 | 1 | 1 | 10 | 15 | 0,666667 |
| 450 | 3,215963 | 0,9 | 0,966667 | 3 | 4 | 0,75 |
| 500 | 4,352871 | 0,933333 | 1 | 17 | 20 | 0,85 |
| 550 | 3,215963 | 1 | 0,866667 | 17 | 19 | 0,894737 |
| 600 | 5,327734 | 0,966667 | 1 | 16 | 20 | 0,8 |
| 650 | 5,327734 | 0,966667 | 1 | 5 | 8 | 0,625 |
| 700 | 4,352871 | 0,966667 | 0,966667 | 4 | 4 | 1 |
| ... | ... | ... | ... | ... | ... | ... |
| 1200 | 5,327734 | 1 | 0,966667 | 8 | 10 | 0,8 |
| 1250 | 7,651052 | 1 | 1 | 16 | 17 | 0,941176 |
| 1300 | 3,215963 | 0,966667 | 0,9 | 8 | 8 | 1 |
| ... | ... | ... | ... | ... | ... | ... |

Аналіз табл. 1 показує, що КФЕ за Кульбаком досягає свого максимального значення $E_{\max} = 7,651052$, яке є граничним для 40 реалізацій за умови відсутності помилок першого та другого роду. Оптимальний крок квантування дорівнює $\tau^* = 400$ градацій, або

$$\tau^* = 400 / dH = 400 / 16000 = 0,025 \text{ с},$$

де dH – початкова частота дискретизації сигналу при оцифруванні акустичного сигналу.

Таким чином, у результаті процесу оптимізації кроку квантування побудовано чіткий в інформаційному сенсі класифікатор, тому що значення критерію досягає свого максимального граничного значення.

ВИСНОВКИ

1. У процесі навчання побудовано безпомилковий за навчальною вибіркою гіперсферичний класифікатор, який забезпечує високу достовірність розпізнавання голосових команд під час алгоритму екзамену.

2. Оптимізація кроку дискретизації підвищує достовірність розпізнавання системи акустичних сигналів.

SUMMARY

OPTIMIZATION OF ACOUSTIC SIGNAL OF QUANTUM STEP AT RECOGNITION OF VOCALS COMMANDS

A.S. Dovbysh, O.A. Yakushev
Sumy State University, Sumy

Optimization of acoustic signal of quantum step is consider 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.

Key words: *IEIT, quantum step, recognition classes, criterion of functional efficiency of Kulbaka, voice signal, work space, algorithm of teaching.*

СПИСОК ЛІТЕРАТУРИ

1. Williams J. and Young S. Partially Observable Markov Decision Processes for Spoken Dialog Systems // Computer Speech and Language. - 2007.- Vol. 21, № 2. – P.231-422.
2. Ронжин А. Л. Система автоматического распознавания русской речи SIRIUS / А. Л. Ронжин, А.А. Карпов, И.В. Ли // Искусственный интеллект. – 2005. – № 3. - С.590-601.
3. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – Киев: Наукова думка, 1987. – 262 с.
4. Касаткина Л.М. Информационная технология быстрого поиска диктора по голосу в базах аудиоданных / Л.М. Касаткина, В.В. Лукович, В.В. Пилипенко // Управляющие системы и машины. - 2006. - № 5, сентябрь-октябрь. – С. 61-67.
5. Краснопопсовський А.С. Інформаційний синтез інтелектуальних систем керування: Підхід, що ґрунтується на методі функціонально-статистичних випробувань.–Суми: Видавництво СумДУ, 2004. – 261 с.
6. Sadaoki Furui, M. Mohan Sondhi Advances in speech signal processing. – New York, 1992. – 871 p.
7. Рабинер Л. Цифровая обработка речевых сигналов / Л. Рабинер, Р. Шафер. - М.: Радио и связь, 1981. - 450 с.
8. Рабинер Л. Теория и применение цифровой обработки сигналов / Л. Рабинер, Б. Гоулд. - М.: Мир, 1976. – 848 с.
9. Нуссбаумер Г. Быстрое преобразование Фурье и алгоритмы вычисления сверток – М.: Радио и связь, 1985. – 248 с.
10. Марпл С.-мл. Цифровой спектральный анализ и его приложения. - М.: Мир, 1990. – 584 с.

Надійшла до редакції 4 квітня 2009 р.