

Метод перетворення мовленого сигналу на основі нерозширювальних рівномірно безперервних відображен

Свєн Е. Федоров, Егле Слесорайтите

Кафедра спеціалізованих комп'ютерних систем
Донецька академія автомобільного транспорту, Донецьк
fee75@mail.ru

Кафедра спеціалізованих комп'ютерних систем
Вільнюський університет, Вільнюс
sliesoraityte@yahoo.com

Анотація

У тезах для вербального управління динамічними об'єктами запропоновано метод перетворення мовленого сигналу, що використовує нерозширювальні рівномірно безперервні відображення.

1. Вступ

На сьогоднішній день широке поширення одержують комп'ютерні системи розпізнавання мовлених образів [1]. Важливу роль для цих систем грає проблема побудови ефективних методів, які забезпечують високу швидкість навчання, високу ймовірність, адекватність і швидкість розпізнавання мови.

Сучасні нейромережеві методи розпізнавання зразків мовленого сигналу використають наступні типи навчанню нейромережевих моделей:

1. Настроювання вагових коефіцієнтів моделі, при цьому опорні зразки відсутні [2]. Недолік – тривале навчання, складність визначення структури моделі, низька ймовірність розпізнавання у випадку важко роздільних класів.

2. Настроювання вагових коефіцієнтів моделі й вибір як опорні зразки тих навчальних зразків, які попадають на границі класів [3]. Недолік – тривале навчання й низька ймовірність розпізнавання у випадку важко роздільних класів.

3. Настроювання вагових коефіцієнтів моделі й формування опорних зразків у вигляді статистичних оцінок безлічі навчальних зразків [4]. Недолік – тривале навчання, складність визначення структури моделі, низька ймовірність розпізнавання у випадку важко роздільних класів.

4. Вибір як опорні зразки всіх навчальних зразків, при цьому вагові коефіцієнти моделі фіксовані [5]. Недолік – триває розпізнавання й низька ймовірність розпізнавання у випадку важко роздільних класів.

5. Формування опорних зразків у вигляді настроювання векторів вагових коефіцієнтів моделі [2]. Недолік – низька ймовірність розпізнавання у випадку важко роздільних класів.

Методи розпізнавання зразків мовного сигналу, засновані на безперервних схованих марковських моделях (СММ) і моделях сумішій розподілів Гаусса (ГСМ), у процесі навчання здійснюють настроювання вагових коефіцієнтів моделі й формування опорних зразків у вигляді статистичних оцінок безлічі навчальних зразків

[6]. Недолік – тривале навчання й низька ймовірність класифікації у випадку важко роздільних класів.

Таким чином, нейромережеві методи й методи, засновані на СММ і ГСМ, мають істотні недоліки.

Метою роботи є створення методу перетворення мовленого сигналу на основі нерозширювальних рівномірно безперервних відображень, що діють у компактних польських просторах.

Для рішення поставленого завдання в тезах конструкуються:

- відображення цифрового сигналу у вектор речовинних ознак;
- відображення вектора речовинних ознак у вектор цілих ознак;
- відображення вектора цілих ознак у номер класу звуку мовлення.

На основі побудованих відображень було проведено чисельне дослідження голосних звуків.

2. Структура нерозширювальних рівномірно безперервних відображень

У роботі [7] були вперше уведені нерозширювальні рівномірно безперервні відображення, що діють у компактних польських просторах зразків сигналів. У даних тезах ці відображення використаються з погляду перетворення й розпізнавання зразків звуків мовлення.

Нерозширювальне рівномірно безперервне відображення $\phi: S^N \rightarrow V^K$, яке діє з компактного польського простору (S^N, ρ) в компактний польський простір (V^K, ρ) , відповідає функції формування вектора ознак, тобто відображає вектор цілих значень дискретного сигналу кінцевої довжини $s = (s_1, \dots, s_N)$, у вектор речовинних ознак $v = (v_1, \dots, v_K)$, причому кожен компонент вектора v обчислюється як логарифмована міра контрастності у вигляді

$$v_{i-l} = \phi_{i-l}(s) = \begin{cases} \lg \left(\frac{E_i}{\sum_{h=l}^i E_h} \right), & E_i > 0 \\ 0, & E_i = 0 \end{cases}, i \in \overline{2, K+1}, (1)$$

$$E_i = \sum_{k=0}^{2^i-1} \left| \sum_{m=0}^{N-1} (s_{m+1} - 2^{r_s}) w_m e^{-j(2\pi/N)km} \right|^2, \quad (2)$$

$$w_m = 0.54 + 0.46 \cos \frac{2\pi m}{N}, \quad (3)$$

де E_i – енергія i -ї октавної смуги Фур'є-спектра,

w_m – вікно Хемминга,

$K+1$ – кількість октавних смуг спектра,

r_s – кількість розрядів (біт) для одного значення сигналу.

Нерозшируване рівномірно безперервне відображення $\psi: V^K \rightarrow X^K$, яку діє з компактного польського простору (V^K, ρ) в компактний польський простір (X^K, ρ) , відповідає функції нормування, масштабування й округлення вектора речовинних ознак, тобто відображає вектор речовинних ознак $\mathbf{v} = (v_1, \dots, v_K)$, у вектор цілих ознак $\mathbf{x} = (x_1, \dots, x_K)$, причому кожна i -я компонента вектора \mathbf{x} обчислюється у вигляді

$$x_i = \psi_i(\mathbf{v}) = \begin{cases} \left[\frac{v_i - \min_{k \in \{1, K\}} v_k}{\max_{k \in \{1, K\}} v_k - \min_{k \in \{1, K\}} v_k} \cdot \alpha \right], & \max_{k \in \{1, K\}} v_k > \min_{k \in \{1, K\}} v_k \\ 0, & \max_{k \in \{1, K\}} v_k = \min_{k \in \{1, K\}} v_k \end{cases}, \quad (4)$$

$$i \in \overline{1, K},$$

де $\alpha \in \{2^{r_X-1}, \dots, 2^{r_X}\}$ – коефіцієнт масштабу, $\lceil \cdot \rceil$ означає округлення.

r_X – кількість розрядів (біт) для одного значення ознаки.

Нерозшируване рівномірно безперервне відображення $\phi: X^K \rightarrow Y^M$, яке діє з компактного польського простору (X^K, ρ) в компактний польський простір (Y^M, ρ) , відповідає функції розпізнавання вектора цілих ознак, тобто відображає вектор цілих ознак $\mathbf{x} = (x_1, \dots, x_K)$ у номер класу зразка звуку мовлення, представлений булевим вектором з одним ненульовим компонентом $\mathbf{y} = (y_1, \dots, y_M)$, причому кожна i -я компонента вектора \mathbf{y} обчислюється у вигляді

$$y_i = \phi_i(\mathbf{x}) = \begin{cases} 1, & \inf_{\tilde{\mathbf{x}} \in X_i^K} \rho(\mathbf{x}, \tilde{\mathbf{x}}) = 0 \\ 0, & \inf_{\tilde{\mathbf{x}} \in X_i^K} \rho(\mathbf{x}, \tilde{\mathbf{x}}) = 1, i \in \overline{1, M}, \end{cases} \quad (5)$$

$$\rho(\mathbf{x}, \tilde{\mathbf{x}}) = \begin{cases} 1, & d(\mathbf{x}, \tilde{\mathbf{x}}) \geq 0 \\ 0, & d(\mathbf{x}, \tilde{\mathbf{x}}) < 0, \end{cases} \quad (6)$$

$$d(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{k=1}^K |\mathbf{x}_k - \tilde{\mathbf{x}}_k|, \quad (7)$$

де X_i^K – клас еквівалентності, що відповідає i -му звуку мовлення.

3. Чисельне дослідження

Області голосних звуків мовлення в ознаковому просторі X^K формуються на основі навчальної множини зразків відповідних звуків. У тезах досліджуються голосні звуки $|i|$, $|y|$, $|o|$, $|a|$, $|ø|$, $|æ|$, не утримуючих переходів ділянок. Параметри для нерозширувальних рівномірно безперервних відображень визначені в роботі [7] у такий спосіб: $K = 7, r_S = 8, r_X = 4, \alpha = 10$.

На рис. 1 наведені навчальні зразки звуку $|i|$ після застосування відображень $\phi: S^N \rightarrow V^K$ й $\psi: V^K \rightarrow X^K$. Область звуку $|i|$ у ознаковому просторі представлена у вигляді $A_i = \{\mathbf{x} | x_1 \in \{10\}, x_2 \in \{4, 5\}, x_3 \in \{4, 5, 6, 7, 8\}, x_4 \in \{7, 8, 9\}, x_5 \in \{2, 3, 4\}, x_6 \in \{0\}, x_7 \in \{1, 2\}\}$.

На рис. 2 наведені навчальні зразки звуку $|y|$ після застосування відображень $\phi: S^N \rightarrow V^K$ й $\psi: V^K \rightarrow X^K$. Область звуку $|y|$ у ознаковому просторі представлена у вигляді $A_y = \{\mathbf{x} | x_1 \in \{10\}, x_2 \in \{6, 7, 8\}, x_3 \in \{5, 6, 7, 8\}, x_4 \in \{5, 6, 7\}, x_5 \in \{3, 4, 5, 6\}, x_6 \in \{0\}, x_7 \in \{1\}\}$.

На рис. 3 наведені навчальні зразки звуку $|o|$ після застосування відображень $\phi: S^N \rightarrow V^K$ й $\psi: V^K \rightarrow X^K$. Область звуку $|o|$ у ознаковому просторі представлена у вигляді $A_o = \{\mathbf{x} | x_1 \in \{10\}, x_2 \in \{6, 7, 8, 9, 10\}, x_3 \in \{6, 7, 8, 9\}, x_4 \in \{1, 2, 3, 4\}, x_5 \in \{1, 2\}, x_6 \in \{0, 1, 2\}, x_7 \in \{0, 1\}\}$.

На рис. 4 наведені навчальні зразки звуку «а» після застосування відображень $\phi: S^N \rightarrow V^K$ й $\psi: V^K \rightarrow X^K$. Область звуку «а» у ознаковому просторі представлена у вигляді $A_a = \{\mathbf{x} | x_1 \in \{10\}, x_2 \in \{4, 5, 6, 7, 8, 9, 10\}, x_3 \in \{4, 5, 6, 7\}, x_4 \in \{1, 2\}, x_5 \in \{1, 2, 3\}, x_6 \in \{1, 2\}, x_7 \in \{0\}\}$.

На рис. 5 наведені навчальні зразки звуку $|e|$ після застосування відображень $\phi: S^N \rightarrow V^K$ й $\psi: V^K \rightarrow X^K$. Область звуку $|e|$ у ознаковому просторі представлена у вигляді $A_e = \{\mathbf{x} | x_1 \in \{10\}, x_2 \in \{3, 4, 5, 6, 7\}, x_3 \in \{2, 3, 4\}, x_4 \in \{5, 6, 7, 8\}, x_5 \in \{1, 2, 3, 4\}, x_6 \in \{0, 1, 2\}, x_7 \in \{0, 1\}\}$.

На рис. 6 наведені навчальні зразки звуку $|i|$ після застосування відображень $\phi: S^N \rightarrow V^K$ й $\psi: V^K \rightarrow X^K$. Область звуку $|i|$ у ознаковому просторі представлена у вигляді $A_u = \{\mathbf{x} | x_1 \in \{10\}, x_2 \in \{4, 5, 6, 7, 8, 9, 10\}, x_3 \in \{3, 4, 5, 6\}, x_4 \in \{5, 6, 7, 8\}, x_5 \in \{2, 3, 4, 5\}, x_6 \in \{0, 1, 2\}, x_7 \in \{0, 1\}\}$.

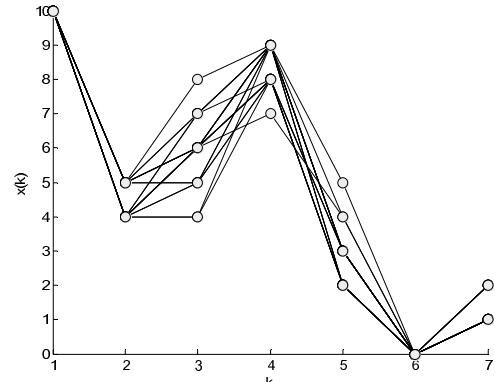


Рисунок 1: Навчальні зразки звуку $|i|$

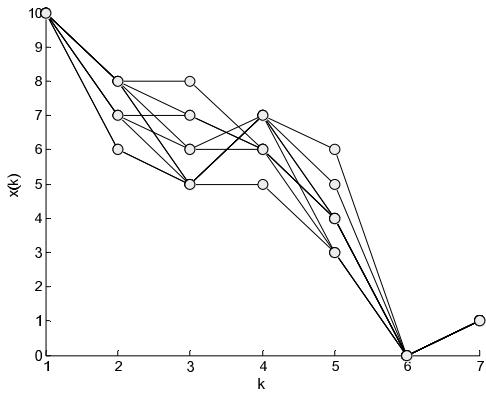


Рисунок 2: Навчальні зразки звука |y|

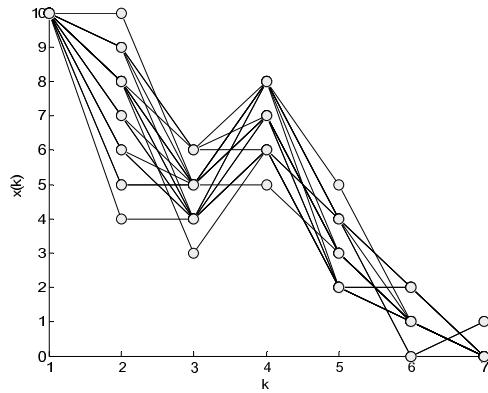


Рисунок 6: Навчальні зразки |i|

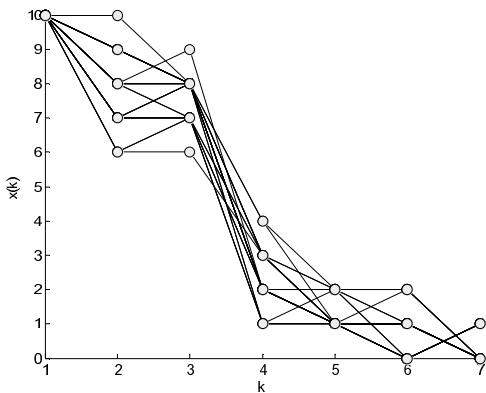


Рисунок 3: Навчальні зразки звука |o|

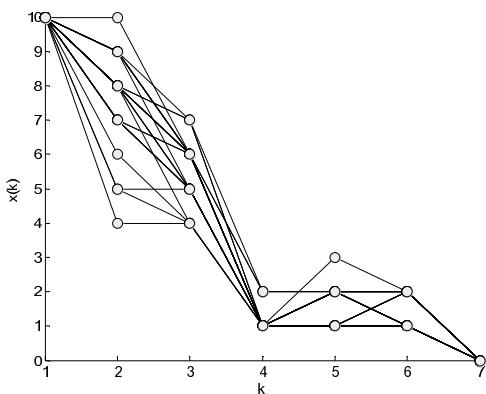


Рисунок 4: Навчальні зразки звука |a|

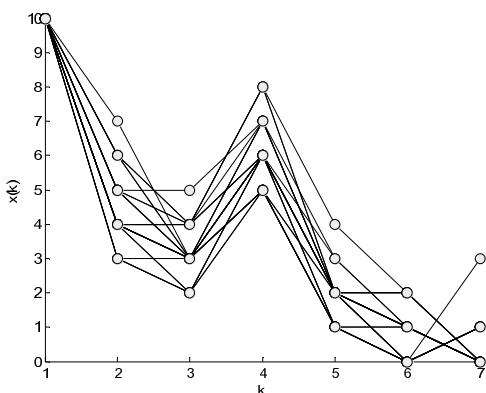


Рисунок 5: Навчальні зразки звука |e|

4. Висновки

У тезах для вербалного керування динамічними об'єктами запропонований метод перетворення й розпізнавання мовних команд оператора на основі нерозширувальних рівномірно безперервних відображень, що діють у компактних польських просторах, що містить у собі: відображення вектора цілих значень дискретного сигналу у вектор речовинних ознак; відображення вектора речовинних ознак у вектор цілих ознак; відображення вектор цілих ознак у номер класу звуку мови. Для запропонованого методу в тезах проведено чисельне дослідження значень векторів цілих ознак голосних звуків і визначені області кожного голосного звуку в ознаковому просторі.

Система ідентифікації оператора по голосу є система розпізнавання мовних команд оператора, розроблені на основі запропонованого методу, можуть використатися системах людино-машинного спілкування для різних галузей (вугільної й металургійної промисловості, авіабудування й суднобудування, нафтопроводів і газопроводів, атомних, теплових і гідроелектростанцій, транспорту й ін.), а також у криміналістиці для фоноскопичної експертизи.

5. Література

- [1] Винцюк, Т.К., *Анализ, распознавание и интерпретация речевых сигналов*, Наук. думка, Київ, 1987, 261 с.
- [2] Осовский, С, *Нейронные сети для обработки информации*, Финансы и статистика, Москва, 2002, 344 с.
- [3] Хайкин, С., *Нейронные сети: полный курс*, Издательский дом «Вильямс», Москва, 2006, 1104 с.
- [4] Комарцова Л.Г., Максимов А.В., *Нейрокомпьютеры*, Изд-во МГТУ им. Н.Э. Батмана, Москва, 2002, 320 с.
- [5] Каллан, Р, *Основные концепции нейронных сетей*, Издательский дом «Вильямс», Москва, 2001, 288 с.
- [6] Rabiner, L.R., Jang, V.H., *Fundamentals of speech recognition*, Prentice Hall PTR, Englewood Cliffs, New Jersey, 1993, 507 р.
- [7] Федоров, Е.Е., “Метод обробки сигналів на основі нерасширяючих рівномірно непреривних отображеній”, *Труды Междунар. науч.-техн. конф. «Радіотехнічні поля, сигнали, апарати та системи (теорія, практика, історія, освіта)» (РТПСАС'2012)*, с. 109-118, 2012.