УДК 681.327.12

А.О. БЕРЕЗА, М.М. БИКОВ, В.В. КОВТУН
Вінницький національний технічний університет

ОПТИМІЗАЦІЯ АЛФАВІТУ ІНФОРМАТИВНИХ ОЗНАК ДЛЯ АВТОМАТИЗОВАНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ МОВЦІВ КРИТИЧНОГО ЗАСТОСУВАННЯ

В результаті проведених досліджень виконано математичну постановку задачі ідентифікації оптимального алфавіту інформативних ознак для застосування у автоматизованих системах розпізнавання мовців критичного застосування. Для цього використано математичний апарат сингулярного аналізу до процесу екстрагування інформативних для розпізнавання мовця ознак із мовного сигналу і сформульовано алгоритм, який дозволяє як ранжувати інформативні ознаки за їх ефективністю для розпізнавання мовця так і, враховуючи комплексний характер інформативних ознак, – оптимізувати розмір їх алфавіту для прийняття рішень в задачі розпізнавання мовців. Запропонований метод базується на математичному перетворенні, функціями якого є власні вектори коваріаційної матриці вхідного сигналу, значення яких дозволяють виділити найменш корельовано інформативні для розпізнавання мовця ознаки. Алгоритм аналізу інформативних ознак на основі запропонованого методу адаптований для систем розпізнавання мовців критичного застосування із блоком прийняття рішень на основі нейромережевого класифікатора глибокого навчання.

Ключові слова: автоматизована система розпізнавання мовців критичного застосування, інформативні ознаки, загортальна нейронна мережа, сингулярний розклад, розклад Карунена-Лоева.

A.O. BEREZA, M.M. BYKOV, V.V. KOVTUN
Vinnytsia National Technical University

INFORMATION FEATURES ALPHABET OPTIMIZATION FOR THE AUTOMATED SPEAKER RECOGNITION SYSTEM OF CRITICAL USE

A research result is performed mathematical formulation of optimal information features alphabet In result of research the mathematical formulation of the problem of informative features optimal alphabet identification for the use in the automated speaker recognition system of critical use. For this purpose the mathematical apparatus was used to the process of extracting of informative for speaker recognition features from speech signal and the algorithm that allows both ranking of informative features by their effectiveness for speaker recognition and, taking into account the complex nature of informative features, to optimize the size of their alphabet for making decisions in speaker recognition task. Offered method is based on mathematical transformation the functions of which are eigenvectors of the covariance matrix of the input signal values of which allow to select the least correlated informative for speaker recognition features. Algorithm of analysis of informative features based on offered method is adapted to speaker recognition systems of critical use with the making decision block on the base of neural network classifier of depth study.

Keywords: automated speaker recognition system of critical use, information features, convolution neural network, a singular transform, Karhunen-Loeve transform.

Вступ

Задача автоматизованого розпізнавання мовця, не дивлячись на свою актуальність, досі не розв'язана, що можна пояснити, в тому числі і її обчислювальною складністю. Існує значна кількість методів аналізу мовних сигналів у частотному та часовому просторах, які дозволяють виділити інформативні для розпізнавання мовця ознаки, але спільною рисою найефективніших з них (наприклад, кепстральні коефіцієнти, період і частота основного тону і т. ін.) є значна обчислювальна складність їх екстрагування та залежність отриманих значень інформативних ознак від якості мовного матеріалу та умов його запису.

В задачах розпізнавання образів знайшли своє застосування методи зменшення розмірності простору ознак, основними з яких є метод Карунена-Лоева [1, 2] (Karhunen-Loeve transform, KLT), метод сингулярного розкладу [3–5] (Singular value decomposition, SVD) та метод головних компонент [6,7] (Principal component analysis, PCA). Спільною рисою цих методів є застосування математичного перетворення, базисними функціями якого є власні вектори коваріаційної матриці вхідного сигналу, з метою виділення найменш корельованих його компонент. В техніці згадані методи в основному використовуються в задачах стиснення та знешумлення зображень та, меншою мірою, знешумлення мовних сигналів. Обмеженість використання цих методів пов'язана із їх значною обчислювальною складністю. Втім у системах автоматизованого розпізнавання останнім часом активно використовують нейромережі глибокого навчання (НМГН), процес навчання яких є складною та трудомісткою операцією. Отже, факторний аналіз інформативних ознак на основі згаданих методів, адаптований для оптимізації алфавіту інформативних ознак для систем розпізнавання мовців нейромережами глибокого навчання, стає ефективним порівняно з НМГН щодо економії обчислювальних ресурсів, що є особливо актуальним для автоматизованих систем критичного застосування.

Постановка завдання

Метою дослідження є отримання математичної постановки задачі ідентифікації оптимального алфавіту інформативних ознак для застосування у автоматизованих систем розпізнавання мовців критичного застосування. Для досягнення поставленої мети автори пропонують застосувати принципи факторного аналізу до процесу екстрагування інформативних для розпізнавання мовця ознак і

сформулювати алгоритм, який дозволить як ранжувати інформативні ознаки за їх ефективністю для розпізнавання мовця так і, враховуючи комплексний характер інформативних ознак, – оптимізувати розмір їх алфавіту для прийняття рішень в задачі розпізнавання мовців.

Математична постановка задачі факторного аналізу інформативних ознак для автоматизованих систем розпізнання мовця критичного застосування

Нехай є M дискретизованих записів мовних сигналів, виголошених одним з m мовців, представлених у просторі інформативних ознак вектором X^m. Для кожного з m мовців відома навчальна вибірка $\left\{X_l^m\right\}$, $l=\overline{1,K}$, де K – обсяг вибірки. Тоді середнє значення по множині записів, що утворюють вибірку для класифікації, $X_{em}^m=\frac{1}{K}\sum_{l=1}^{K}X_l^m$ буде вважатися оцінкою належності аналізованого сигналу до одного з m мовців.

Використовуючи можливості сингулярного розкладу можна отримати оператор розкладу $B^m=\left[B_1^m,B_2^m,...,B_N^m\right]^T$ для класів m, якщо для рядків B_i^m матриці B^m виконується умова ортонормованості.

Виконаємо екстрагування інформативних для розпізнавання мовців ознак з вектора вихідних даних X^m:

$$Y^m=\frac{1}{N}B^mX^m, \tag{1}$$

де Y^m – вектор інформативних для розпізнавання мовців коефіцієнтів, перетворений за системою ортогональних функцій B^m для класу m.

Для кожного вектора X^m, який належить класу m векторів вихідних даних, отримаємо $Y^m=B^mX^m$, де $X^m=\left\lfloor x_1^m,x_2^m,...,x_N^m\right\rfloor$, $Y^m=\left\lfloor y_1^m,y_2^m,...,y_N^m\right\rfloor$, тоді процес обчислення X^m можна представити відношенням

$$X^m=\sum_{i=1}^{N}y_i^mB_i^m. \tag{2}$$

Мінімізуючи обсяг алфавіту інформативних ознак необхідно не втратити адекватності опису класів мовців представлених вектором вихідних даних X^m. Використаємо середньоквадратичний критерій для оцінювання адекватності перетворення. Отримаємо оцінку $\widetilde{X}^m$ вектора X^m, представивши його M членами, замінивши решту $N\times M$ членів y_i^m константами c_i^m:

$$\widetilde{X}^m=\sum_{i=1}^{M}y_i^mB_i^m+\sum_{i=1}^{N}c_i^mB_i^m. \tag{3}$$

Похибка опису вектора X^m його оцінкою $\widetilde{X}^m$ описується вектором похибки

$$\Delta X^m=X^m-\widetilde{X}^m. \tag{4}$$

Тоді середньоквадратична похибка матиме вид:

$$\sigma^m=\sum_{i=M+1}^{N}B_i^{mT}K_x^mB_i^m, \tag{5}$$

де K_x^m – коваріаційна матриця.

Інформативність коефіцієнтів ознак, виділених із мовного сигналу, при розкладі вектора вихідних даних X^m в ортогональному базисі (1) можна оцінити відповідним значенням дисперсії $\left(\sigma_i^m\right)^2$ і узагальнити у вигляді характеристики

$$\Phi_r^m\left(Y^m\right)=\frac{\sum_{i=1}^{p}\left(y_i^m\right)^2}{\sum_{i=1}^{p}\left(\sigma_i^m\right)^2}, \tag{6}$$

де $\left(\sigma_i^m\right)^2$ – дисперсія коефіцієнтів інформативних ознак. Характеристику $\Phi_r^m\left(Y^m\right)$ називатимемо оптимальним розкладом. При певному виборі розмірності підпростору інформативних ознак r

запропонована характеристика $\Phi_r^m\left(Y^m\right)$ має екстремальні властивості, що дозволяє використовувати її в тому числі і для класифікації. Якщо вектор X^q розкладається в базисі B^m класу m, то такий розклад буде найбільш точним при $q=m$ і $X^q \in \left\{X_i^m\right\}$. Характеристика $\Phi_r^m\left(Y^m\right)$ при цьому буде максимальною. Дане твердження справедливе для всіх векторів навчальної вибірки, тому характеристику $\Phi_r^m\left(Y^m\right)$ можна використана для формулювання критерію класифікації, що враховує оптимальність представлення вектора вихідних даних: якщо $\Phi^{mq}=\Phi_r^m-\Phi_r^q>0$ для всіх $q=\overline{1,M}$, $(q\neq m)$, то $X\in\left\{X^m\right\}$.

Для ефективної класифікації із використанням запропонованого критерію необхідно, щоб характеристика Φ^{mq} була максимальною за всіма векторами навчальної вибірки і по всіх парах різних класів. Для вибору розмірності підпросторів p_M сформулюємо функцію для максимізації Φ^{mq}:

$$F\left(r_j\right)=\frac{1}{M}\sum_{m=1}^{M}\frac{1}{M-1}\sum_{\substack{q=1\\(q\neq m)}}^{M}\Phi^{mq}\rightarrow\max,\ j=\overline{1,M}\ . \tag{7}$$

Оптимальні розмірності підпросторів r_M можна отримати, підставивши у функцію (7) рівняння для Φ^{mq} і розв'язавши отримані задачі максимізації. Нехай розклад векторів X^q за всіма базисами B^q, окрім базису, побудованого для векторів q-го класу, матиме рівномірний розподіл. Тоді математичне очікування коефіцієнтів інформативних ознак буде наближатися до рівномірного, що робить припустимим таке співвідношення:

$$E\left(\left(y_i^q\right)^2\right)=\frac{\left\|Y^q\right\|}{N}\ . \tag{8}$$

Підставимо значення математичного очікування (8) у (7), вважаючи при цьому, що норма вектора вихідних даних $\left\|Y^q\right\|=1$. В результаті функція (7) набуватиме виду

$$F\left(r_j\right)=\frac{1}{M}\sum_{m=1}^{M}\frac{1}{M-1}\sum_{\substack{q=1\\(q\neq m)}}^{M}\left(\sum_{i=1}^{r_s}\frac{\left(y_i^m\right)^2}{\left(\sigma_i^m\right)^2}-\frac{r_S}{N}\sum_{i=1}^{r_s}\frac{1}{\left(\sigma_i^m\right)^2}\right),\ j=\overline{1,M},. \tag{9}$$

Отже, задача максимізації функції (9) представляється множиною задач максимізації функцій однієї змінної:

$$F_q\left(r_q\right)=\sum_{i=1}^{r_q}\frac{\left(y_i^m\right)^2}{\left(\sigma_i^m\right)^2}-\frac{r_q}{N}\sum_{i=1}^{r_q}\frac{1}{\left(\sigma_i^m\right)^2}\ . \tag{10}$$

Значення розмірності підпросторів, при яких функції (10) набувають своїх максимальних значень, можна оцінити урахуванням умови (8) і припущенням, що розклад векторів навчальної вибірки, які належать m-му класу, у базисі B^m, буде відповідати умові (6), і буде виконується умова $\left(y_1^m\right)^2\geq\left(y_2^m\right)^2\geq\ldots\geq\left(y_N^m\right)^2\geq 0$. Приріст для функцій $F_q\left(r_q\right)$ із зростанням розмірності перестає бути позитивним, якщо виконується умова $\left(y_i^m\right)^2\geq\frac{1}{N}\geq\left(y_{i+1}^m\right)^2$.

Множина функцій (10) має екстремальну властивість, що дозволяє використовувати їх в якості критеріїв класифікації дискретних сигналів за коефіцієнтами розкладів у базисах, субоптимальних за сингулярним розкладом. Запропонований критерій не враховує значень $N-r$ коефіцієнтів інформативних ознак.

Алгоритм факторного аналізу інформативних ознак для автоматизованих систем розпізнавання мовців критичного застосування

Узагальнимо запропоновані теоретичні дослідження застосування сингулярного розкладу для оптимізації алфавіту інформативних ознак для розпізнавання мовців у вигляді алгоритму, адаптованого до застосування у автоматизованих системах розпізнавання мовців критичного застосування в умовах ризику (статистичної невизначеності):

1. Задаємо мінімально допустиму середньоквадратичну помилку класифікації мовців, обумовлену виключенням частини ознак із вихідної їх множини – $\left(\sigma_i^m\right)^2$;

2. Центруємо векторні вибіркові дані вимірювань інформативних ознак;

3. Обчислюємо вибіркову коваріаційну матрицю K_x^m для центрованих векторних вибіркових даних $X^m = \left\lfloor x_1^m, x_2^m, ..., x_N^m \right\rfloor$:

$$K_x = \frac{1}{N}\sum_{i=1}^{N} X_i^m X_i^{mT};$$

4. Обчислюємо власні значення p_i і власні вектори $V_i, i = \overline{1,N}$ коваріаційної матриці K_x^m:

$$K_x V_i = p_i V_j,\ i,\ j = \overline{1,N},$$

де

$$\left(V_i, V_j\right) = \begin{cases} a_i,\ якщо\ i = j, \\ 0,\ якщо\ i \neq j; \end{cases}$$

5. Нормуємо власні вектори;

6. Впорядковуємо власні значення за спаданням;

7. Виключаємо з подальшого розгляду власні значення з індексами $i = \overline{m+1,N}$ згідно з заданою середньоквадратичною помилкою

$$\left(\sigma_i^m\right)^2 = \sum_{i=m+1}^{N} p_i;$$

8. Формуємо базис із власних векторів, які відповідають власним значенням $p_1, p_2, ..., p_m$, які залишилися;

9. Обчислюємо компоненти Y_i, $i = \overline{1,m}$, в сформованому базисі із власних векторів V_i:

$$Y_i = V_i X,\ i = \overline{1,m}.$$

Отримані в результаті роботи алгоритму результати є розв'язком задачі вибору оптимального алфавіту інформативних ознак для класифікації мовців.

Експериментальні дослідження

Проведемо обчислювальний експеримент застосування запропонованого алгоритму оптимізації алфавіту барк-кепстрального аналізу записів мовних сигналів, використовуючи який модуль прийняття рішень автоматизованої системи розпізнавання мовців критичного використання здійснюватиме процес розпізнавання.

В якості бази еталонних записів використано безкоштовну базу даних NOIZEUS [8] – спеціалізовану базу даних Школи інжинірингу та комп'ютерних наук Еріка Джонсона при Університеті Техасу в Далласі, США, яка використовується для дослідження алгоритмів покращення звуку і складається з 30 речень англійської розмовної мови, вимовлених трьома чоловіками та трьома жінками (по 5 на кожного мовця, частота дискретизації записів складає 25 кГц, але задля додавання шуму була зменшена до 8 кГц) та записів типових побутових та техногенних шумів. В ході експерименту систему автоматизованого розпізнавання мовців критичного застосування навчали як записами чистих парольних фраз, так і парольними фразами із додаванням шуму. В першому випадку навчальна вибірка містила 18 парольних фраз, в другому – 576, де до чистого сигналу додавався штучний шум з рівнями шум/сигнал 0 дБ, 5 дБ, 10 дБ, 15 дБ відповідно.

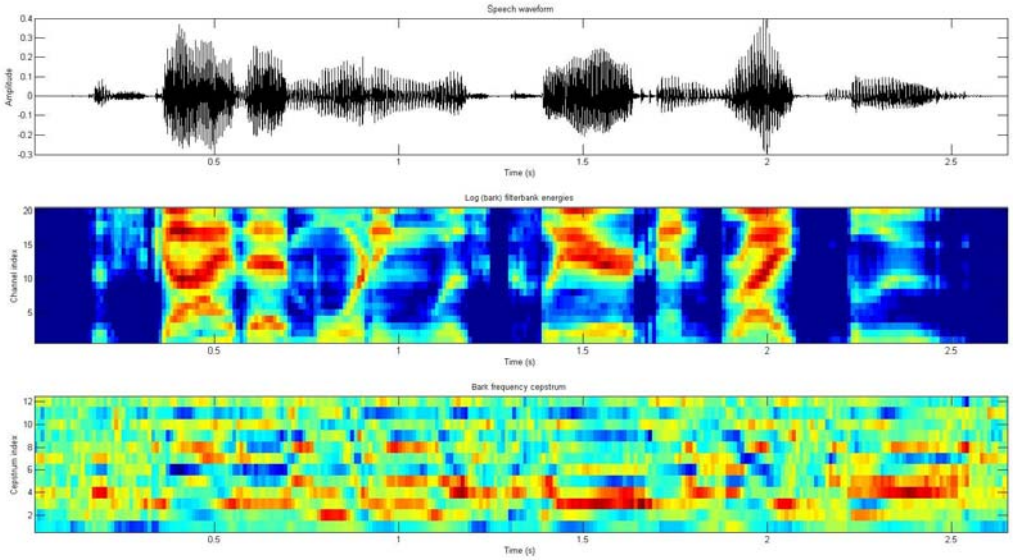

Рис. 1. Приклад роботи модуля виділення барк-кепстральних коефіцієнтів мовного сигналу

Для отримання кепстральних коефіцієнтів вхідний сигнал поділявся на кадри, тривалістю 20 мс, з кожного з яких виділялися 12 кепстральних коефіцієнтів, 12 дельта-коефіцієнтів (перша похідна) і 12 подвійних дельта кепстральних коефіцієнтів (друга похідна). Гребінка 20 трикутних смугових фільтрів перекривала частотний діапазон 40-8000 Гц, координати точок фільтрів визначалися так, щоб кожна пара фільтрів перекривалася на половину і на викривленій частотній шкалі кожен фільтр починається і закінчується в центрі сусідньої фільтру. Приклад роботи модуля виділення барк-кепстральних коефіцієнтів з мовного сигналу наведено на рисунку 1.

В результаті роботи запропонованого алгоритму оптимізації алфавіту інформативної ознаки проводиться селекція частотних смуг за принципом мінімізації коефіцієнтів матриці міжсмугової кореляції, що дозволяє відсіяти менш інформативні смуги і проводити навчання класифікатору на максимально інформативному матеріалі. Для розпізнання мовців за оптимізованим алфавітом інформативної ознаки використано згортальну нейромережу [9], архітектура якої наведена на рисунку 2. Практичну реалізацію загортальної нейромережі глибокого навчання виконано засобами кросплатформних бібліотек Caffe [10] із відкритим програмним кодом. На розпізнавання мовців нейромережу було навчено із використанням алгоритму стохастичного градієнтного спуску (Stochastic Gradient Descent Algorithm) [11].

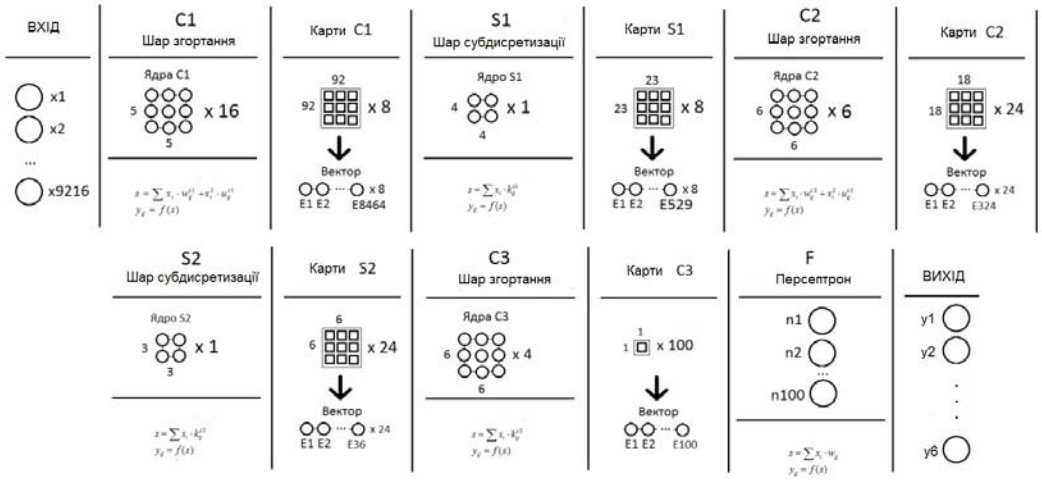

Рис. 2. Архітектура згортальної нейромережі глибокого навчання для розпізнавання мовців за оптимізованими інформативними ознаками

Для навчання нейромережі використано 60% обсягу бази аудіозаписів, в яку увійшли екземпляри записів без шумів та із різним рівнем шум/сигнал (5, 10. 15 дБ) відповідно. Тестуюча вибірка склала решту 40% аудіозаписів. Узагальнені результати експерименту представлено на рисунку 3, де імовірність правильного розпізнавання розраховувалася за формулою

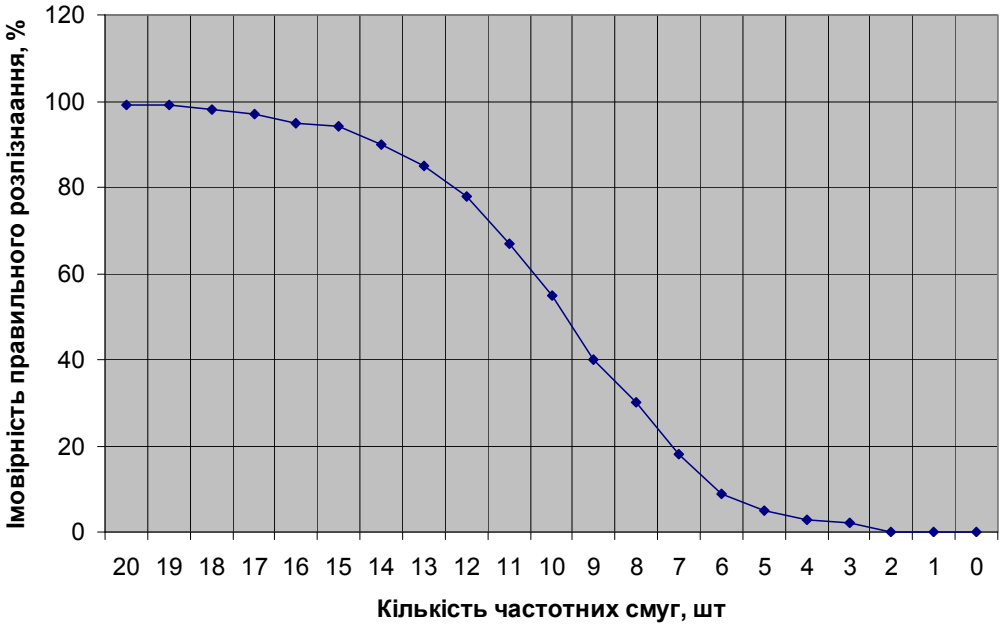

Рис. 3. Залежність імовірності правильного розпізнавання мовців від кількості частотних смуг, інформативні ознаки з яких використовувалися для розпізнавання

$$P_{pr} = \frac{\sum_i (Np_i)}{N}, \quad (11)$$

де Np_i – кількість правильних результатів розпізнавання i-го мовця, N – загальна кількість експериментів.

Як видно з рисунка, імовірність правильного розпізнавання мовців перевищувала 90% при використанні для розпізнавання 14 частотних смуг із початкових 20 і перевищувала 95% при кількості частотних смуг рівній 16. Така інформація дозволяє зменшити обсяг навчальної вибірки на 30% або 20% відповідно із прогнозованою втратою імовірності правильного розпізнавання. Такою виявилося, що визначені алгоритмом як менш інформативні для розпізнавання мовців є частотні смуги розташовувалися в 4200-8000 Гц.

Висновки

Отже, авторами здійснено математичну постановку задачі ідентифікації оптимального алфавіту інформативних ознак для застосування у автоматизованих систем розпізнавання мовців критичного застосування. Для цього автори застосували принципи сингулярного аналізу до процесу екстрагування інформативних для розпізнавання мовця ознак і сформулювати алгоритм, який дозволить як ранжувати інформативні ознаки за їх ефективністю для розпізнавання мовця так і, враховуючи комплексний характер інформативних ознак, – оптимізувати розмір їх алфавіту для прийняття рішень в задачі розпізнавання мовців.

Експериментальні дослідження довели працездатність запропонованого алгоритму на прикладі оптимізації алфавіту барк-кепстру мовних сигналів, коефіцієнти якого використовувалися для розпізнавання мовців загортальною нейромережею. Застосування алгоритму дозволило виявити залежність між імовірністю правильного розпізнавання мовця і кількістю частих смуг, в яких виділялася інформативна ознака, а також визначити менш інформативні частотні смуги.

Література

1. Dony R.D. Optimally adaptive transform coding / Dony R.D., Haykin S. // IEEE Trans. Image Processing, 1995. – 4(10). – P. 1358–1370.

2. Moulin P. The role of linear semi-infinite programming in signal-adapted QMF bank design / Moulin P., Anitescu M., Kortanek K.O., Potra F. // IEEE Trans. Signal Processing. – 1997. – Vol. 45. – P. 2160–2174.

3. Abdallah S. A. Application of geometric dependency analysis to the separation of convolved mixtures / Abdallah S. A., Plumbley M. D. // Proc. of the International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2004), Granada, Spain, 2004. – P. 22–24.

4. Jafari M. G. Sparse coding for convolutive blind audio separation / Jafari M. G., Abdallah S. A., Plumbley M. D., Davies M. E. – Springer-Verlag, Berlin. Proc. ICA, 2006. – P. 132–139,

5. Michal A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representations / Michal A., Michael E., Bruckstein A. // IEEE Trans. on Signal Processing. – 2006. – Vol. 54. – P. 4311–4322.

6. Gorban A. N. Principal Manifolds for Data Visualisation and Dimension Reduction, Series: Lecture Notes in Computational Science and Engineering 58, Springer, Berlin / Gorban A. N., Kegl B., Wunsch D., Zinovyev A. Y. – Heidelberg. – New York, 2007, XXIV. – 340 p. – ISBN 978-3-540-73749-0.

7. Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed. – Springer, NY, 2002. – 487 p. – ISBN 978-0-387-95442-4

8. NOIZEUS: Noisy speech corpus - Univ. Texas-Dallas [Електронний ресурс]. – Режим доступу : http://ecs.utdallas.edu/loizou/speech/noizeus/.

9. CS231n: Convolutional Neural Networks for Visual Recognition [Електронний ресурс]. – Режим доступу : http://cs231n.github.io/convolutional-networks/

10. Caffe | Deep Learning Framework [Електронний ресурс]. – Режим доступу : http://caffe.berkeleyvision.org/.

11. An overview of gradient descent optimization algorithms [Електронний ресурс]. – Режим доступу : http://sebastianruder.com/optimizing-gradient-descent/.

Отримана/Received : 8.5.2017 р. Надрукована/Printed :10.6.2017 р.
Рецензент: д.т.н., проф.Бісікало О.В.