

УДК 004.588.032.26:004.89:004.932.2

**В. В. МОСКАЛЕНКО, А. С. МОСКАЛЕНКО, А. Г. КОРОБОВ**

*Сумський державний університет, Україна*

## **МЕТОД НАВЧАННЯ БЕЗ ВЧИТЕЛЯ ІЄРАРХІЧНОГО ЕКСТРАКТОРА ВІЗУАЛЬНИХ ОЗНАК НА ОСНОВІ МОДИФІКАЦІЇ НЕЙРОННОГО ГАЗУ**

*Розроблено алгоритм навчання без учителя згорткової мережі для екстракції візуальних ознак на основі принципів нейронного газу та розрідженого кодування. Пропонується метод інформаційно-екстремального машинного навчання, оснований на використанні двійкового кодування ознакового подання спостережень та популяційного алгоритму оптимізації параметрів функціонування системи розпізнавання для формування вирішальних правил та оцінки ефективності навченого екстрактора ознак. Порівнюється за інформаційним критерієм, що є функціоналом від помилок першого та другого роду, ефективність навчання вирішальних правил у випадку використання екстрактора, навченого з учителем методом стохастичного градієнтного спуску, та у випадку екстрактора, навченого без учителя.*

**Ключові слова:** *нейронний газ, згорткова нейронна мережа, розріджене кодування, інформаційний критерій, система контрольних допусків, ознаки розпізнавання.*

### **Вступ**

Технології машинного аналізу візуальної інформації набули широкого застосування в робототехнічних системах та інфокомунікаційних сервісах різного призначення. Основними рушіями розвитку машинного інтелекту в системах аналізу візуальних образів є прогрес в мікропроцесорній техніці та технологіях навчання глибоких нейронних мереж. Проте сучасні методи інтелектуальної обробки даних потребують значних обчислювальних ресурсів та обсягів навчальних даних, що ускладнює їх впровадження в автономних системах з обмеженими ресурсами. Крім того, існує брак робочих алгоритмів неперервного (само) навчання глибоких моделей даних в режимі реального темпу часу, що обмежує ефективність їх функціонування за умов нестаціонарності та апріорної невизначеності.

Одним з підходів до зменшення обчислювальної складності глибокого машинного навчання є так звана передача знань (Transfer Learning) [1]. В рамках даного підходу можна використати акумульовані в іншій задачі знання щодо аналізу візуальних образів і при необхідності виконати редукцію чи донавчання моделі даних для її адаптації під нову задачу. При цьому задача неперервного машинного донавчання ієрархічного ознакового подання без учителя за новими спостереженнями в невідомому середовищі може бути вирішена за рахунок використання автоенкодерів чи обмеженої машини Больцмана, проте ці моделі є повнозв'язними мережами і тому досить ресурсоемкими [2]. У працях [3, 4] пропонується використовувати принципи нейронного газу для навчання словника базисних векторів, що

використовуються для розрідженого кодування ознакового опису спостережень. При цьому було доведено швидшу збіжність даного алгоритму самонавчання ознак і зниження ефекту перенавчання (overfitting) вирішальних правил при обмеженому обсязі навчальної вибірки. Однак досі не було розглянуто використання даного підходу для самонавчання ієрархічного екстрактора ознак, наприклад, оснований на згортковій нейронній мережі.

Основним компонентом інтелектуальної системи аналізу даних є вирішальні правила, що часто представляють із себе класифікатор ознакового опису спостережень. Найбільш популярним алгоритмом класифікаційного аналізу є метод опорних векторів, проте цей алгоритм потребує багато ручних налаштувань для регуляризації моделі і його продуктивність залежить від складності ядерних функцій перетворення простору ознак [1]. Перспективним підходом є використання двійкового кодування спостережень і радіально-базисних функцій в бінарному просторі Хеммінга на основі операцій порівняння та “виключаючого АБО”, оскільки ці операції є найбільш обчислювально прості [5, 6]. При цьому використання інформаційних критеріїв ефективності та популяційних алгоритмів пошуку оптимальних параметрів функціонування дозволяє реалізувати автоматичну регуляризацію моделі даних.

В даній роботі пропонується використати принципи нейронного газу та розрідженого кодування для навчання ієрархічного екстрактора візуальних ознак. При цьому оцінку ефективності екстрактора передбачається здійснювати за результатами навчання інформаційно-екстремального класифікатора з двійковим кодуванням спостережень.

## 1. Постановка задачі

Розглянемо формалізовану постановку задачі інформаційного синтезу здатного до самонавчання ієрархічного екстрактора візуальних ознак. Дано модель відомої згорткової мережі VGG-16 [7], конфігурацію десяти шарів якої показано на рис. 1.



Рис. 1. Конфігурація перших десяти шарів згорткової нейронної мережі VGG-16 [7, 8]

Дано навчальну вибірку із бази зображень ImageNet [8], за якою сформовано вибірку патчів зображення  $\{x_i^{(j)} \mid i = \overline{1, N}, j = \overline{1, n}\}$ , де  $N, n$  – кількість ознак розпізнавання ( $N = 9$ ) та патчів відповідно.

Необхідно в процесі машинного навчання без учителя визначити вагові коефіцієнти першого згорткового шару у вигляді  $M$  базисних векторів  $C = \{c_{m,i} \mid m = \overline{1, M}, i = \overline{1, N}\}$ . На основі отриманих карт ознак за випадково обраними входними зображеннями потрібно сформувати нову вибірку патчів наступного шару і продовжувати пошарове самонавчання до десятого шару включно.

Необхідно порівняти ефективність розпізнавання навчальної вибірки зображень ImageNet при використанні в екстракторі ознак, отриманих в процесі навчання з учителем, з ознаками, сформованими шляхом навчання без учителя, за максимальним значенням інформаційного критерію

$$\bar{E}^* = \frac{1}{K} \sum_{k=1}^K \max_{\{s\}} E_k, \quad (1)$$

де  $E_k$  – інформаційний критерій навчання вирішальних правил розпізнавати реалізації класу  $X_k^0$ ;

$K$  – кількість класів розпізнавання;

$\{s\}$  – множина кроків машинного навчання.

## 2. Алгоритм навчання без вчителя ієрархічного екстрактора ознак

Важливим кроком аналізу зображень є їх попередня нормалізація з метою усунення лінійної кореляції компонент спостереження і уніфікації первинного ознакового подання. Відбілювання даних за

методом ZCA (Zero-phase Component Analysis) є одним з найпоширеніших методів попередньої нормалізації даних. Метод ZCA полягає у виконанні наступних кроків:

1) обчислення середнього вибіркового значення яскравості пікселів  $X$  в патчах  $\mu := \text{mean}(X)$ ;

2) обчислення коваріаційної матриці вибірко-вих спостережень  $\Sigma := \text{cov}(X)$ ;

3) сингулярний розклад коваріаційної матриці  $\Sigma \approx VDV^T$ ;

4) відбілювання кожного спостереження за формулою  $x_j := VD^{-\frac{1}{2}}V^T(x_j - \mu)$ .

Навчання фільтрів згорткового шару пропонується здійснювати на основі модифікації нейронного газу, запропонованої в працях [3, 4]. Цей алгоритм має такі основні кроки:

1) ініціалізація словника базисних векторів  $C$  випадковими значеннями з рівномірного розподілу;

2) вибір  $t$ -го входного спостереження  $x$  з множини  $X$ , яка має обсяг  $t_{\max}$ ;

3) нормалізація базисних векторів  $C$  шляхом їх приведення до одиничного розміру (unit vector);

4) обчислення коефіцієнтів розміру околу сусідства та швидкості навчання за формулами:

$$\lambda_t := \lambda_o (\lambda_{\text{final}} / \lambda_o)^{t/t_{\max}}, \quad \alpha_t := \alpha_o (\alpha_{\text{final}} / \alpha_o)^{t/t_{\max}},$$

де  $\lambda_o, \lambda_{\text{final}}$  – початкове та кінцеве значення коефіцієнту  $\lambda_t$ ;  $\alpha_o, \alpha_{\text{final}}$  – початкове та кінцеве значення коефіцієнту  $\alpha_t$ ;

5) ініціалізація множини індексів тих стовпців матриці  $C$ , які вже були використані протягом  $t$ -ітерацій,  $U = \emptyset$ ;

6) ініціалізація залишку, що мінімізується,  $\varepsilon^U := x$ ;

7) ініціалізація тимчасової матриці  $R := (\tau_1, \dots, \tau_1, \dots, \tau_M) = C$ , ортонормованої відповідно до  $C^U$ ;

8) ініціалізація лічильника кроків уточнення залишків  $h := 1, h = \overline{1, K-1}$ ;

9) розрахунок міри схожості вектора  $\tau_k$  до  $\varepsilon^U$  та впорядкування їх за зростанням

$$-(\tau_0^T \varepsilon^U)^2 \leq \dots \leq -(\tau_k^T \varepsilon^U)^2 \leq \dots \leq -(\tau_{M-h-1}^T \varepsilon^U)^2$$

ініціалізація лічильника кроків уточнення базисних векторів  $C$ ,  $k := 0, k = \overline{0, M-h-1}$ ;

10) оновлення на  $k$ -му кроці базисних векторів з використанням принципів ортогональності до підпростору заданого в  $C^U$  і правила Ойа [4]

$$c_{lk} := c_{lk} + \Delta_{lk}, \quad \tau_k := \tau_k + \Delta_{lk},$$

де  $\Delta_{lk} := \alpha_t \exp(-k/\lambda_t) y(\varepsilon^U - \eta_{lk}^U)$ , де  $y := r_{lk}^T \varepsilon^U$ ;

11) нормалізація  $r_{lk}$  шляхом приведення до одиничного розміру (unit vector);

12) якщо  $k < 0, M - h - 1$  перехід до кроку 11;

13) визначення базису переможця за формулою  $l_{win} := \arg \max_{l, l \in U} (r_l^T \varepsilon_i^U)^2$ ;

14) оновлення матриці  $R$  та поточного залишку  $\varepsilon_i^U$  за формулами:

$$r_l := r_l - (r_{l_{win}}^T r_l) r_{l_{win}}, \quad \varepsilon_i^U := \varepsilon_i^U - (r_{l_{win}}^T \varepsilon_i^U) r_{l_{win}},$$

де  $r_{l_{win}}$  – стовбчик матриці  $R$ , що має максимальне перекриття з поточним залишком  $\varepsilon_j^U$ , індекс якого ще не додано до  $U$ ;

15) оновлення матриці обраних базисних векторів  $U = U \cup l_{win}$ ;

16) якщо  $h < K - 1$ , то перехід до кроку 11;

17) якщо  $t < t_{max}$ , то перехід до кроку 2, інакше – закінчення обробки.

При побудові інформаційно-екстремальних вирішальних правил двійкове кодування ознакового опису відбувається шляхом порівняння значення  $i - \bar{i}$  ознаки з відповідним нижнім  $A_{B,l,i}$  та верхнім  $A_{T,l,i}$  межами несиметричного поля контрольних допусків, які розраховуються за формулами

$$A_{B,l,i} = y_{i,max} \left[ 1 - \frac{\delta_{l,i}}{\delta_{max}} \right],$$

$$A_{T,l,i} = y_{i,max}, \text{ при } l = \overline{1, L},$$

де  $\delta_{l,i}$  – параметр  $l$ -го поля контрольних допусків на значення  $i - \bar{i}$  ознаки.

Формування бінарної навчальної матриці  $\{b_{k,i}^{(j)} \mid i = \overline{1, L} \cdot N; j = \overline{1, n_k}; k = \overline{1, K}\}$ , де  $N$  – кількість ознак класифікатора,  $n_k$  – кількість векторів класу  $X_k^0$  та  $K$  – кількість класів розпізнавання, здійснюється за правилом

$$b_{k, (l-1) \cdot N + i}^{(j)} = \begin{cases} 1, & \text{if } A_{B,l,i} \leq y_{k,i}^{(j)} \leq A_{T,l,i}; \\ 0, & \text{else.} \end{cases}$$

Обчислення значень координат двійкового етапного (усередненого) вектора  $b_k$ , відносно якого відбувається побудова в радіальному базисі контейнерів класів, здійснюється за правилом

$$b_{k,i} = \begin{cases} 1, & \text{if } \frac{1}{n_k} \sum_{j=1}^{n_k} b_{k,i}^{(j)} > \frac{1}{n} \sum_{k=1}^K \sum_{j=1}^{n_k} b_{k,i}^{(j)}; \\ 0, & \text{if else;} \end{cases} \quad i = \overline{1, N \cdot L},$$

де  $n$  – повний обсяг розмічених векторів навчальної вибірки.

Як критерій ефективності машинного навчання класифікатора розпізнавати реалізації класу  $X_k^0$  розглядається модифікація інформаційної міри Кульбака [5, 6]:

$$J_k = \frac{1 - (\alpha_k + \beta_k)}{\log_2(2 + \varepsilon) - \log_2 \varepsilon} \cdot \log_2 \left[ \frac{2 - (\alpha_k + \beta_k) + \varepsilon}{(\alpha_k + \beta_k) + \varepsilon} \right], \quad (2)$$

де  $\alpha_k, \beta_k$  – оцінки помилок першого та другого роду, які задають робочу область критерію у вигляді нерівностей  $\alpha_k \geq 0,5$  та  $\beta_k \geq 0,5$ ;  $\varepsilon$  – мале знакододатне число для уникнення невизначеності при діленні на нуль, рівне, як правило, числу з діапазону  $[10^{-4} \dots 10^{-2}]$ .

Оптимізація параметрів поля контрольних допусків  $\{\delta_{l,i}\}$  полягає в пошуку екстремуму функції критерію (2) в гіперпросторі рішень. При цьому як пошуковий алгоритм в даній роботі пропонується використати рій частинок (Particle Swarm Optimization, PSO), який характеризується простотою реалізації та інтерпретабельністю [6]. Оптимізація радіусів контейнерів класів може здійснюватися методом послідовного прямого перебору з заданим кроком, оскільки кількість кроків такого пошуку є відносно малою.

Таким чином, алгоритм навчання без учителя ієрархічного екстрактора ознак полягає в кластер-аналізі патчів зображення на основі модифікації алгоритму нейронного газу на кожному з шарів згорткової мережі для формування відповідного алфавіту згорткових фільтрів. При цьому машинне навчання вирішальних правил дозволяє оперативно отримати оцінку ефективності в інформаційному розумінні отриманого ознакового подання.

### 3. Результати фізичного моделювання

Запропонований алгоритм машинного навчання без учителя багатшарової згорткової нейронної мережі VGG-16 був використаний для синтезу екстрактора ознакового опису та класифікатора об'єктів з алфавіту  $\{X_k^0 \mid k = \overline{1, K}\}$ . Оцінку ефективності екстрактора ознак пропонується здійснювати шляхом навчання класифікатора розрізняти класи з такого алфавіту ( $K=21$ ): "фон", "літак", "велосипед", "птаха", "човен", "пляшка", "автобус", "машина", "кішка", "стілець", "корова", "обідній стіл", "собака", "колячка", "мотоцикл", "людина", "горщик рослини", "вівця", "диван", "поїзд", "монітор". Обсяг навчальної та тестової вибірок для кожного класу

становить  $n_k = 100$  зображень. Для навчання екстрактора було випадковим чином сформовано вибірку патчів  $3 \times 3$  з зображень бази ImageNet. Обсяг вибірки патчів становить  $n = 1000000$ .

На рис. 2 показано навчений банк фільтрів для кожного з каналів зображення для першого шару мережі, приведеного на рис. 1.

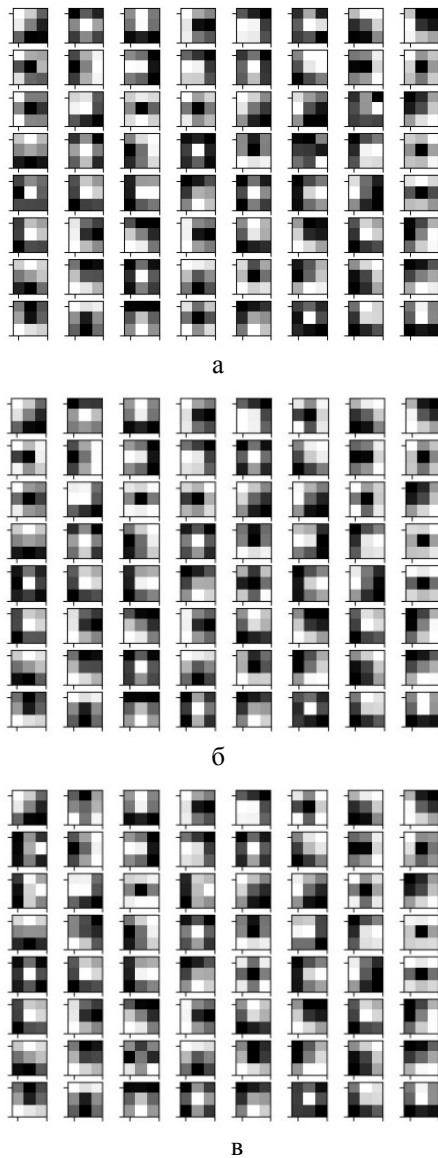


Рис. 2. Зображення згорткових фільтрів першого шару VGG-16: а – канал червоного кольору; б – канал зеленого кольору; в – канал синього кольору

Навчання без вчителя за запропонованим алгоритмом виконано для десяти шарів згорткової мережі. На основі сформованого ознакового подання було здійснено навчання класифікатора. Графіки зміни максимумів інформаційного критерію (1) в процесі ройової оптимізації параметра поля контрольних допусків  $\delta_{1,i}$  для ознакового подання, сформованого 10-ма шарами екстрактора, як

навченого на повному обсязі вибірки ImageNet методом стохастичного градієнтного спуску [8], так і навченого без учителя за модифікованим алгоритмом нейронного газу, показано на рис. 3. Лічильник кроків  $k$  відповідає кількості міграцій агентів рою.

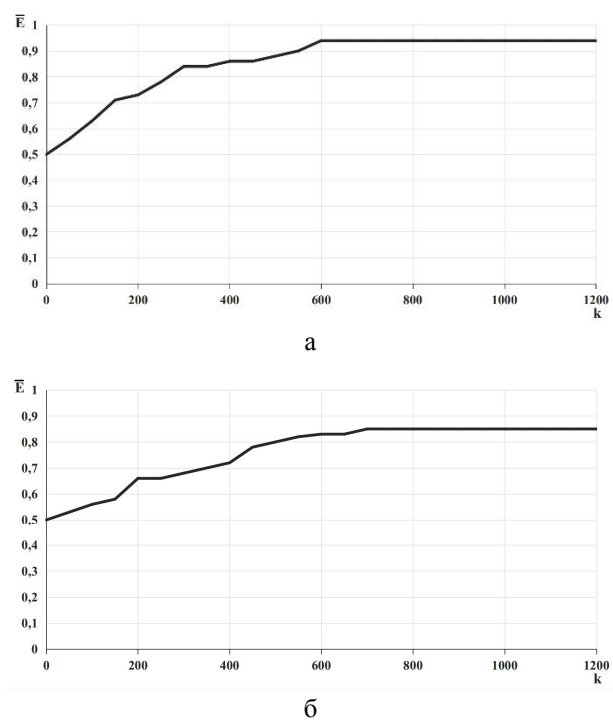


Рис. 3. Графік зміни максимумів усередненого критерію (1) в процесі оптимізації вектора  $\{\delta_i\}$ : а – екстрактор ознак, навчений з учителем; б – екстрактор ознак, навчений без учителя

Аналіз рис. 3. показує, що протягом 1200 ітерацій ройового алгоритму пошуку знайдено глобальний оптимум функції критерію (1). При цьому значення критерію (1) для екстрактора, навченого з учителем, рівне  $\bar{E}^* = 0,81$ , що відповідає вибірковій точності  $A = 0,987$ . Значення критерію (1) для екстрактора, навченого за запропонованим алгоритмом без учителя, рівне  $\bar{E}^* = 0,75$ , що відповідає вибірковій точності  $A = 0,964$ .

Таким чином, запропонований алгоритм машинного навчання екстрактора ознак без учителя дозволяє здійснювати синтез вирішальних правил з прийнятною точністю для багатьох практичних застосувань, але дещо поступається результатам, отриманим при навчанні екстрактора з учителем.

## Висновки

1. Доведено за результатами фізичного моделювання ефективність навчання без вчителя ієрархічного екстрактора ознак, побудованого на перших

десяти шарах згорткової нейронної мережі VGG-16, шляхом використання модифікації алгоритму нейронного газу, спрямованої на розріджене кодування спостережень.

2. Показано, що використання навчання з учителем за алгоритмом стохастичного градієнтного спуску дозволяє отримати для алфавіту з 21 класу розпізнавання точність вирішальних правил рівну 98,7%, в той час, як запропонований алгоритм навчання екстрактора ознак без вчителя дещо поступається і забезпечує точність рівну 96,4%.

3. Перспективним напрямком подальших досліджень є розробка алгоритмів метаоптимізації для уточнення фільтрів екстрактора ознак та параметрів алгоритму його навчання без вчителя.

Робота підтримана в рамках науково-дослідної роботи ДР № 0117U003934 "Інтелектуальна автономна бортова система безпілотного літального апарату для ідентифікації об'єктів на місцевості".

## Література

1. Zhongling, H. *Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data [Text]* / H. Zhongling, P. Zongxu, L. Bin // *Remote Sensing*. – 2017. – T. 9, No 909. – P. 2–21.
2. *Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction [Text]* / J. Masci, U. Meier, D. Cirezan, J. Schmidhuber // *Proceedings of the International Conference on Artificial Neural Networks (ICANN 2011), Part I, Espoo, Finland, 14-17th June, 2011*. – Springer-Verlag Berlin Heidelberg, 2011. – P. 52–59.
3. Labusch, K. *Learning Data Representations with Sparse Coding Neural Gas [Text]* / K. Labusch, E. Barth, Th. Martinetz // *Proceedings of the European Symposium on Artificial Neural Networks, Bruges, Belgium, 23–25 April, 2008*. – d-side publications. – P. 233–238.
4. Labusch, K. *Sparse Coding Neural Gas: Learning of Overcomplete Data Representations [Text]* / K. Labusch, E. Barth, T. Martinetz // *Neurocomputing*. – 2009. – Vol. 72, I. 7–9. – P. 1547–1555.
5. Moskalenko, V. V. *Information Extreme Method for Classification of observations with categorical attributes [Text]* / A. S. Rizhova, V. V. Moskalenko, A. S. Dovbysh // *Cybernetics and Systems Analysis*. – 2016. – Vol. 52, No 2. – P. 56–63.
6. Moskalenko, V. V. *Optimizing the parameters of functioning of the system of management of data center it infrastructure [Text]* / V. V. Moskalenko, S. V. Pimonenko // *Eastern-european journal of enterprise technologies*. – 2016. – Vol. 5, No 2 (83). – P. 21–29.
7. *Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning [Text]* / H.-W. Ng, V. Dung Nguyen, V. Vonikakis, S. Winkler // *Proceedings of the International Conference On Multimodal Interaction (ICMI'15), Seattle, WA, USA, 09–13 November, 2015*. – P. 443–449.
8. Simonyan, K. *Very Deep Convolutional Networks for Large-Scale Image Recognition [Text]* / K. Simonyan, A. Zisserman // *Proceedings of the 3rd International Conference on Learning Representations (ICLR2015), Hilton San Diego Resort & Spa, 7–9 May, 2015*. – P. 1–14.

## References

1. Zhongling, H., Zongxu, P. and Bin, L. *Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data. Remote Sensing*, 2017, vol. 9, no. 909, pp. 2–21. Available at: <http://www.mdpi.com/2072-4292/9/9/907/pdf>
2. Masci, J., Meier, U., Cirezan, D., Schmidhuber, J. *Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. Proc. International Conference on Artificial Neural Networks (ICANN 2011)*, 2011, pp. 52–59. DOI:10.1007/978-3-642-21735-7\_7.
3. Labusch, K., Barth, E., Martinetz, Th. *Learning Data Representations with Sparse Coding Neural Gas. Proc. 16th European Symposium on Artificial Neural Networks*, 2008, pp. 233–238. DOI: 10.1007/978-3-540-87536-9\_81.
4. Labusch, K., Barth, E., Martinetz, Th. *Sparse Coding Neural Gas: Learning of Overcomplete Data Representations. Neurocomputing*, 2009, vol. 72, i. 7–9, pp. 1547–1555. DOI:10.1016/j.neucom.2008.11.027.
5. Rizhova, A.S., Moskalenko, V. V., Dovbysh, A. S. *Information Extreme Method for Classification of observations with categorical attributes. Cybernetics and Systems Analysis*, 2016, vol.52, no. 2, pp. 56–63. DOI:10.1007/s1055.
6. Moskalenko, V. V., Pimonenko, S. V. *Optimizing the parameters of functioning of the system of management of data center it infrastructure. Eastern-European journal of enterprise technologies*, 2016, vol. 5, no. 2 (83), pp. 21–29. DOI: 10.15587/1729-4061.2016.79231.
7. Ng, H.-W., Dung Nguyen, V., Vonikakis, V., Winkler, S. *Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning. Proc. 17th International Conference on Multimodal Interaction (ICMI'15)*, 2015, pp. 443–449. DOI:10.1145/2818346.2830593.
8. Simonyan, K., Zisserman, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition. Proc. 3rd International Conference on Learning Representations (ICLR2015)*, 2015, pp. 1–14.

Надійшла до редколегії 18.08.2017 розглянута на редколегії 14.09.2017

## МЕТОД ОБУЧЕНИЯ БЕЗ УЧИТЕЛЯ ИЕРАРХИЧЕСКОГО ЭКСТРАКТОРА ВИЗУАЛЬНЫХ ПРИЗНАКОВ НА ОСНОВЕ МОДИФИКАЦИИ НЕЙРОННОГО ГАЗА

*В. В. Москаленко, А. С. Москаленко, А. Г. Коробов*

Разработан алгоритм обучения без учителя свёрточной сети для экстракции визуальных признаков на основе принципов нейронного газа и разреженного кодирования. Предлагается метод информационно-экстремального машинного обучения, основанный на использовании двоичного кодирования признакового представления наблюдений и популяционного алгоритма оптимизации параметров функционирования системы распознавания для формирования решающих правил и оценки эффективности обученного экстрактора признаков. Сравнивается по информационному критерию, который является функционалом от ошибок первого и второго рода, эффективность обучения решающих правил в случае использования экстрактора, обученного с учителем методом стохастического градиентного спуска, и в случае экстрактора, обученного без учителя.

**Ключевые слова:** нейронный газ, сверточная нейронная сеть, разреженное кодирование, информационный критерий, система контрольных допусков, признаки распознавания.

## METHOD OF UNSUPERVISED LEARNING OF HIERARCHICAL EXTRACTOR OF VISUAL FEATURES BASED ON MODIFICATION OF NEURAL GAS

*V. V. Moskalenko, A. S. Moskalenko, A. G. Korobov*

The modern technologies of the intellectual analysis of visual information for solving the problem of unsupervised training in real time with the aim of adapting to unknown conditions of observation are analyzed. It is proposed to use 10 layers of the well-known neural network VGG-16 as a model of the hierarchical extractor of visual features that can be used in the transfer learning tasks. The use of the principles of the neural gas to increase the convergence rate of the algorithm of unsupervised learning of the extractor of visual features under the conditions of a limited amount of training data is considered. The modification of the neuron gas aimed to sparse coding of input observations is based on the optimized orthogonal matching pursuit algorithm that was used to increase the informativeness of the feature set in condition of limited sample size. Training dataset is generated by selecting from a popular image base ImageNet and selecting patches from selected images or feature maps on a given layer.

The method of so-called information-extreme machine learning of decision rules is proposed for assessing the efficiency of the proposed feature extractor. Information-extreme learning is based on the use of binary coding of the feature representation of observations and the construction of radial-basic decision rules in Hamming's binary space. The implementation of the algorithm is based on the use of computationally simple operations such comparison with threshold and a bitwise XOR. Optimization of the geometric parameters of the partition feature space into separated classes is carried out in the binary space, therefore, it can be implemented by the method of a sequential direct busting with a given step, since such steps are relatively small. For optimizing parameters of encoding observations rules is used population-based particle swarm algorithm for searching global maximum of logarithmic information Kullback's criterion in admissible domain of it function. In this case we normalized modification information criterion which is function of the first and second kind errors is used.

The effectiveness of training of decision rules in the case of the use of an extractor supervised trained with by a stochastic gradient descent method, with case of supervised trained feature extractor is compared. According to the results of physical modeling unsupervised learning of extractor ensures the accuracy of decisive rules to 96.4% which is inferior to the accuracy of supervised learning which is equal to 98.7% are shown.

**Keywords:** neural gas, convolution neural network, sparse coding, information criterion, the system of control tolerances, features of recognition pattern.

**Москаленко Вячеслав Васильович** – канд. техн. наук, доцент каф. комп'ютерних наук, Сумський державний університет, Суми, e-mail: v.moskalenko@cs.sumdu.edu.ua.

**Москаленко Альона Сергіївна** – канд. техн. наук, асистент каф. комп'ютерних наук, Сумський державний університет, Суми, e-mail: a.moskalenko@cs.sumdu.edu.ua, alenarizhova@gmail.com.

**Коробов Артем Геннадійович** – аспірант каф. комп'ютерних наук, Сумський державний університет, Суми, e-mail: artemkorr@gmail.com.

**Moskalenko Viacheslav Vasilyovich** – Ph.D., associate professor of Computer Sciences Department of Sumy State University, Sumy, e-mail: v.moskalenko@cs.sumdu.edu.ua.

**Moskalenko Alyona Sergiyvna** – Ph.D., teaching assistant of Computer Sciences Department of Sumy State University, Sumy, e-mail: a.moskalenko@cs.sumdu.edu.ua, alenarizhova@gmail.com.

**Korobov Artem Gennadiyovich** – Ph.D.-student of Computer Sciences Department of Sumy State University, Sumy, e-mail: artemkorr@gmail.com.