

УДК 004.89

**М.П. Комар, канд. техн. наук, Р.М. Перевізник, Д.Б. Неспляк, Р.Є. Комарницький,
Т.М. Червоняк, В.Р. Вигнанець, В.Р. Деньчук, О.М. Голодюк, Д.В. Гатенюк**
Західноукраїнський національний університет, Україна

**ПРОЕКТУВАННЯ ПРИКЛАДНИХ СИСТЕМ ОБРОБКИ ТА АНАЛІЗУ
ВЕЛИКИХ ДАНИХ НА ОСНОВІ ГЛИБОКИХ НЕЙРОННИХ МЕРЕЖ**

**M.P. Komar, Ph.D, R.M. Pereviznyk, D.B. Nespliak, R.Ye. Komarnytskyi,
T.M. Chervoniak, V.R. Vyhnanets, V.R. Denchuk, O.M. Holodiuk, D.V. Hateniuk**
**DESIGN OF APPLIED SYSTEMS OF BIG DATA PROCESSING AND ANALYSIS
BASED ON DEEP NEURAL NETWORKS**

Аналіз робіт зарубіжних вчених показав, що великий і швидко зростаючий обсяг цифрової інформації вимагає розробки новітніх технологій її обробки та аналізу [1]. Проблеми обробки та аналізу великих даних, що пов'язані з їхнім різноманіттям, з труднощами збору, зберігання, управління та аналізу, об'ємом пам'яті та швидкістю обчислень наведені в [2-4], описані методики та алгоритми, що використовуються для управління великими наборами даних.

Для вирішення проблем зберігання й доступу до великих даних необхідна розподілена кластерна платформа. Така система має забезпечувати великий простір для зберігання (петабайт) і прозорий доступ до файлів даних на сервери кластера.

Сьогодні методи машинного навчання разом з досягненнями в області обчислювальної потужності стали відігравати життєво важливу роль в обробці та аналізі великих даних [5-8]. Глибокі нейронні мережі [9-15] останніми роками набули широкого поширення, істотно витіснивши більшість інших методів у таких галузях: машинне навчання, комп'ютерний зір і опрацювання сигналів. Підставами для такого стрімкого успіху глибоких нейронних мереж є два чинники: наявність великих об'ємів даних і здешевлення обчислювальних потужностей. Проте крім очевидних переваг у вигляді поліпшення і розширення функціоналу комп'ютерних систем, реалізація алгоритмів глибокого навчання на стандартних ПК спричиняє очевидні труднощі, пов'язані із нестачею ресурсів, насамперед, пам'яті й обчислювальних потужностей процесора.

Основними недоліками відомих рішень є: залежність від стрімкого зростання обсягу даних у багатьох предметних галузях, відсутність методів аналізу різнотипних даних, необхідність у значних людських ресурсах для підтримки процесу аналізу даних, висока обчислювальна складність наявних алгоритмів аналізу даних призводять до постійного зростання часу аналізу даних навіть при регулярному оновленні апаратних засобів; необхідність роботи із розподіленими базами даних, можливості яких більшість існуючих методів аналізу даних не використовують ефективно.

Можливість спільного використання штучних нейронних мереж і високопродуктивних систем істотно розширює область практичних задач та є актуальним напрямком на шляху усунення вищезазначених недоліків.

Отже, новизна представленого матеріалу полягає в розробці методів збору та зберігання великих даних, інтелектуальної обробки та аналізу великих даних в різних предметних областях на основі глибоких нейронних мереж, паралельного навчання глибоких нейронних мереж, забезпечення стійкості до вторгнень кіберфізичних систем.

Підвищення продуктивності інтелектуальної обробки та аналізу великих даних планується досягти за рахунок використання методів паралельного навчання глибоких нейронних мереж та методів зменшення розмірності великих даних. Паралельне

навчання глибоких нейронних мереж дозволить збільшити швидкість навчання та зменшити використання пам'яті графічних процесорів. Проведені дослідження показали, чим більша розмірність мережі, тим істотніший вигравш у часі при використанні паралельного підходу. Це відбувається тому, що час виконання послідовної частини коду, яка міститься в паралельній реалізації, при збільшенні навчальної вибірки, практично, не збільшується, тоді як час виконання повністю послідовної реалізації зростає пропорційно.

Для виявлення та класифікації вторгнень на кіберфізичні системи запропоновано використати методи штучних нейронних мереж. У зв'язку із здатністю штучних нейронних мереж в процесі навчання виявляти складні залежності між вхідними і вихідними даними, які були відсутні в навчальній вибірці, і здатністю коректно класифікувати зашумлені образи, вони є привабливим інструментом для вирішення складних різноманітних задач захисту інформації.

Література

1. X.-W. Chen and X. Lin. Big Data Deep Learning: Challenges and Perspectives. IEEE Access, Vol.2, pp. 514–525, 2014.
2. Shirkhorshidi A. S. Big data clustering: a review / A. S. Shirkhorshidi et al. // International Conference on Computational Science and Its Applications. – Springer, Cham, 2014. – P. 707-720.
3. Marjani M. Big IoT data analytics: architecture, opportunities, and open research challenges / M. Marjani et al. // IEEE Access. – 2017. – Т. 5. – P. 5247-5261.
4. Kurasova O. Strategies for big data clustering/ O. Kurasova et al. // 2014 IEEE 26th International Conference on Tools with Artificial Intelligence. – IEEE, 2014. – P. 740-747.
5. J. Lin, A. Kolcz. Large-scale machine learning at twitter // Proc. ACM SIGMOD Scottsdale Arizona USA, pp. 793-804, 2012.
6. B. Panda at al. MapReduce and its application to massively parallel learning of decision tree ensembles // Scaling Up Machine Learning: Parallel and Distributed Approaches, 2012.
7. Krizhevsky A at al. ImageNet classification with deep convolutional neural networks // In Proc. Advances in Neural information Processing Systems. – 2012. – 25. – P. 1090–1098.
8. Hinton G. at al. Deep neural network for acoustic modeling in speech recognition / G. Hinton // IEEE Signal Processing Magazine. – 2012. – № 29. – P. 82–97.
9. Li Deng and Dong Yu. Deep Learning: Methods and Applications. Foundations and Trends in Signal Processing, Vol. 7, N. 3-4, pp. 197–387, 2014.
10. J. Schmidhuber. Deep learning in neural networks: An overview // Neural Networks, Vol. 61, N. 1, pp. 85–117, January 2015.
11. Fischer and C. Igel. Training Restricted Boltzmann Machines: An Introduction. Pattern Recognition, Vol. 47, N. 1, pp. 25–39, January 2014.
12. Y. Bengio at al. Representation Learning: A Review and New Perspectives // IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, N. 8, pp. 1798–1828, August 2013.
13. LeCun, Y., Bengio, Y., Hinton, G. Deep learning // Nature. – 2015. – 521 (7553). – P. 436–444.
14. Golovko, V. The Nature of Unsupervised Learning in Deep Neural Networks: A New Understanding and Novel Approach / Vladimir Golovko, Aliaksandr Kroshchanka, Douglas Treadwell // Optical Memory and Neural Networks (Springer Link). – 2016. – Vol. 25, № 3. – P. 127–141.
15. Golovko, V.A. Deep learning: an overview and main paradigms // Optical Memory and Neural Networks (Information Optics). – 2017. – Vol. 26, № 1. – P. 1–17.